Ruggedness and Interlaboratory Studies for Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test: Phase I Report

U.S. Department of Transportation
Federal Highway Administration

FOREWORD

The standards for Asphalt Mixture Performance Tester (AMPT) cyclic fatigue testing (AASHTO TP 107-18 and AASHTO TP 133-19) enable pavement engineers to predict asphalt-mixture performance over a wide range of loading and climate conditions.[1,2] These standards are an important component of ongoing Federal Highway Administration efforts to increase pavement life through fundamental testing and mechanistically-based predictive models. To enable widespread implementation of the procedures for AMPT cyclic fatigue testing, tolerance ranges for the experimental factors that contribute to variability in the test results must be defined using sound statistical methods. This report details a ruggedness evaluation of the standards for AMPT cyclic fatigue testing that was undertaken to meet this need. The results of the ruggedness-evaluation experiments were analyzed to isolate the effect of the individual factors and establish statistically justified limits on the experimental factors to limit variability in the test results. These limits, combined with several supplementary studies, were used to establish recommendations to improve both AASHTO TP 107-18 and TP 133-19. Draft revised versions of both standards have also been developed as part of this effort.

Cheryl Allen Richter, P.E., Ph.D.
Director, Office of Infrastructure
Research and Development

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. FHWA-HRT-21-057	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Ruggedness and Interlaboratory Studies for Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test: Phase I Report		**5. Report Date** May 2021
		6. Performing Organization Code
7. Author(s) Cassie Castorena (ORCID: 0000-0002-5915-0084), B. Shane Underwood (ORCID: 0000-0002-7223-3968), Y. Richard Kim (ORCID: 0000-0003-3295-977X), Kangjin Lee (ORCID: 0000-0003-3458-0747), Nam Tran (ORCID: 0000-0001-8183-4741), and Adam Taylor		**8. Performing Organization Report No.**
9. Performing Organization Name and Address North Carolina State University Department of Civil, Construction, and Environmental Engineering Campus Box 7908 Raleigh, NC 27695		**10. Work Unit No.**
		11. Contract or Grant No. DTFH6117C00037
12. Sponsoring Agency Name and Address Office of Infrastructure Research and Development Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101		**13. Type of Report and Period Covered** Phase I Report; September 2017–April 2020
		14. Sponsoring Agency Code HRDI-10

15. Supplementary Notes
The Contracting Officer's Representative is David J. Mensching, Ph.D., P.E. (HRDI-10).

16. Abstract
This report highlights findings from Phase I of a project designed to evaluate the ruggedness of the standards for cyclic fatigue testing using an Asphalt Mixture Performance Tester (AMPT). These standards include AASHTO TP 107-18 and 133-19, which are applicable to 100-mm diameter and 38-mm diameter test specimens, respectively.[1,2] The ruggedness evaluation conducted adheres to the requirements of ASTM E1169-14 and integrates ASTM C1067-12 recommendations,[3,4] which are specifically for construction materials. The experimental factors evaluated included air-void content, specimen height, diameter difference between loading platen and specimen, loading-platen parallelism, rest period between the dynamic modulus fingerprint and cyclic fatigue test, strain level, and temperature. Based on the outcomes of this experiment and several supplementary studies, changes to the AASHTO TP 107-18 and TP 133-19 protocols were proposed.[1,2]

17. Key Words Cyclic fatigue testing, Asphalt Mixture Performance Tester, AMPT, asphalt mixture, ruggedness evaluation, asphalt mixture performance testing standards.	**18. Distribution Statement** No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. http://www.ntis.gov		
19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages** 130	**22. Price** N/A

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
LENGTH				
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
AREA				
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
VOLUME				
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: volumes greater than 1,000 L shall be shown in m^3				
MASS				
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2,000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
TEMPERATURE (exact degrees)				
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
ILLUMINATION				
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
FORCE and PRESSURE or STRESS				
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
LENGTH				
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
AREA				
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
VOLUME				
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
MASS				
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2,000 lb)	T
TEMPERATURE (exact degrees)				
°C	Celsius	1.8C+32	Fahrenheit	°F
ILLUMINATION				
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m2	0.2919	foot-Lamberts	fl
FORCE and PRESSURE or STRESS				
N	newtons	2.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations

2S2P1D	two springs, two parabolic elements, one dashpot
AASHTO	American Association of State Highway and Transportation Officials
AC	asphalt concrete
AMPT	Asphalt Mixture Performance Tester
ASTM	American Society for Testing and Materials
DOT	department of transportation
ESAL	equivalent-single-axle load
FHWA	Federal Highway Administration
IDT	indirect tension
ILS	interlaboratory study
KEC	Korea Expressway Corp.
LVDT	linear variable differential transformer
MIT-RAP	Manitoba Infrastructure and Transportation Reclaimed Asphalt Pavement
MnROAD	Minnesota roadway test track
NCAT	National Center for Asphalt Technology
NMAS	nominal maximum aggregate size
PG	performance grade
PID	proportional, integral, derivative
PRS	Performance-Related Specifications
RAP	reclaimed asphalt pavement
S-VECD	simplified viscoelastic continuum damage

Symbols

a_1	fitting coefficient for exponential damage-characteristic curve model
a_2	fitting coefficient for exponential damage-characteristic curve model
α_T	time–temperature shift factor at a given temperature
α	damage growth rate
b	sigmoidal model fitting coefficient
B_{0-7}	threshold regression model coefficients
β	2S2P1D fitting coefficient
C	pseudostiffness
C_{11}	fitting coefficient for power law damage-characteristic curve model
C_{12}	fitting coefficient for power law damage-characteristic curve model
C_i	pseudosecant modulus at the current time step
C_{i-1}	pseudosecant modulus at the previous time step
C^*	cyclic pseudosecant modulus
C^*_n	cyclic pseudosecant modulus at the current analysis cycle
C^*_{n-1}	cyclic pseudosecant modulus at the previous analysis cycle
C_h	specific heat (W-s/kg-°C)
%Cracking	ratio of the sum of the fatigue-cracking area to the total lane area
d	sigmoidal model fitting coefficient

D^R	pseudostrain energy-based fatigue failure criterion
DMR	dynamic modulus ratio
%Damage	amount of damage in the cross section of the asphalt layer
δ	2S2P1D fitting coefficient
$E'(\omega,T)$	storage modulus at a particular temperature and angular frequency (kPa or psi)
$E'(\omega_R)$	storage modulus at a particular reduced angular frequency (kPa or psi)
E'_{2S2P1D}	storage modulus from 2S2P1D model
$E'_{Sigmoidal}$	storage modulus from the Sigmoidal model
E_0	maximum storage modulus value (kPa or psi)
$|E^*|$	dynamic modulus at 10 Hz and given temperature (kPa)
$|E^*|_{fingerprint}$	dynamic modulus determined from fingerprint test
$|E^*|_{LVE}$	dynamic modulus from 2S2P1D model
$e(t)$	error
Error	threshold regression model error
ε^R	pseudostrain
ε^R_{pp}	peak-to-peak pseudostrain
ε^R_{ta}	tension amplitude of the pseudostrain
g	sigmoidal model fitting coefficient
g^e	effective gap (mm)
g^m	measured gap (mm)
G^R	average rate of pseudostrain energy release
γ	2S2P1D fitting coefficient
h	2S2P1D fitting coefficient
k_c	thermal conductivity (W/m-°C)
K_1	form adjustment factor
K_d	derivative gain
K_i	integral gain
K_p	proportional gain
κ	2S2P1D fitting coefficient
λ	convective heat coefficient (W/m²/°C)
λ_{AMPT}	AMPT convective coefficient
λ_{lab}	convective flow coefficient
max E'	maximum storage modulus
n	total number of data points collected during test
N	number of cycles
N_f	number of cycles to failure
ω	angular frequency (rad/s).
ω_R	reduced angular frequency (rad/s)
φ_{platen}	loading-platen diameter (mm)
$\varphi_{specimen}$	test-specimen diameter (mm)
q	loading-platen angle (degree)
r	radius (m)
ρ	density (kg/m³)
S	damage parameter
S_{app}	apparent damage capacity
SE	standard error of actuator strain

ΔS_i	damage growth between the current and previous time step
ΔS_n	damage growth between the current and previous analysis cycle
σ	stress (kPa or psi)
σ_{pp}	peak-to-peak stress (kPa or psi)
t	time measured from the experiment (s)
t_R	reduced time (s)
T	temperature (°C)
T_{ext}	external temperature (°C)
$T_{x,t}$	current temperature at position x and time t (°C)
τ_E	2S2P1D fitting coefficient
$u(t)$	output
x_i	measured actuator strain at point i
$\hat{x}_i$	predicted actuator strain at point i from the sinusoid
$\hat{x}_0$	amplitude of the sinusoid
$X_{1\text{-}7}$	seven factors in the ruggedness test
Y	predicted *%Cracking*

CHAPTER 1. INTRODUCTION

The American Association of State Highway and Transportation Officials (AASHTO) testing protocols designated AASHTO TP 107-18 and 133-19, collectively referred to as procedures for Asphalt Mixture Performance Tester (AMPT) cyclic fatigue testing, enable the practical, mechanistic performance characterization of asphalt mixtures using cyclic fatigue testing in the AMPT.[1,2] These test procedures are a key component in the Federal Highway Administration (FHWA) Performance-Related Specifications (PRS) initiative. They are also the only testing protocol to assess fatigue cracking that has been developed specifically for the AMPT.

AASHTO TP 107-18 applies to cylindrical asphalt mixture specimens that are 100-mm diameter by 130-mm tall (hereafter referred to as large specimens). In contrast, AASHTO TP 133-19 applies to cylindrical asphalt mixture specimens that are 38-mm diameter by 110-mm tall (hereafter referred to as small specimens). AASHTO TP 107-18 preceded AASHTO TP 133-19. The small-specimen geometry was established to improve testing efficiency and enable field core testing for mixtures with nominal maximum aggregate size (NMAS) values up to 19 mm.

The results of AMPT cyclic fatigue testing coupled with the results of dynamic modulus ($|E^*|$) testing are used to obtain two key material functions of the Simplified Viscoelastic Continuum Damage (S-VECD) model: the damage-characteristic curve (i.e., psuedostiffness (C) versus the internal state parameter representing damage (S)) and the pseudostrain energy-based fatigue failure criterion (D^R). The damage-characteristic curve and failure criterion are independent of mode of loading, loading history, and temperature. Consequently, prediction of the damage response to any given loading history of interest is possible using limited test results, thus making the fatigue-cracking characterization of asphalt mixtures efficient compared to empirical methods.

Although the mechanistic approach described is advantageous, the standards for AMPT cyclic fatigue testing need to include tolerance ranges for the experimental factors that contribute to variability in the test results. The objective of this study was to address this shortcoming and improve the standards for AMPT cyclic fatigue testing based on a ruggedness evaluation. To achieve this objective, researchers from two laboratories (designated lab A and lab B in this study) conducted a ruggedness experiment and, based on the outcomes of this experiment, recommended changes to the AASHTO TP 107-18 and TP 133-19 protocols. A ruggedness evaluation involves performing repeated experiments using an established or provisional test protocol but making systematic changes in specific variables, also referred to as factors. The evaluation conducted in this study adhered to the requirements of ASTM E1169-14 and integrated ASTM C1067-12 recommendations, which are specifically for construction materials.[3,4] This integration aligns with the approaches followed in the ruggedness evaluations conducted for AASHTO T 378-17 (then AASHTO TP 79)[5,6] and AASHTO TP 131-18[7] for indirect tension (IDT) dynamic modulus testing.[8]

The results of these experiments were observed and analyzed to isolate the effect of each individual factor and, ultimately, used to establish statistically justified limits on the factors to better control repeatability and reproducibility of the protocol. In this study, the research team

conducted the ruggedness evaluation prior to performing an interlaboratory inquiry that will establish the repeatability and reproducibility of the standards.

This report is divided into four chapters and an appendix. Chapter 1 introduces the project, the test methods, the objectives, and the study. Chapter 2 provides the overall methodology of the study, including the materials, test and analysis methods, and supplementary experiments carried out. Chapter 3 presents the results of the experiments and statistical analysis. This chapter is presented in as concise a manner as possible to facilitate easier review of the main results of the experimental study and statistical analysis. However, detailed results of all experiments are presented in the appendix. Finally, chapter 4 presents the conclusions and recommended changes to the test standards.

This report uses certain naming conventions throughout: the term "component materials" refers to asphalt binder and aggregate from a hot-mix asphalt plant, and the terms "test specimens" or "specimens" refer to laboratory-mixed materials that have been laboratory-compacted, cored, and cut to the final dimensions used for testing.

CHAPTER 2. METHODOLOGY

ASTM E1169-14 and C1067-12 describe the basic methodology for carrying out the ruggedness evaluation.[3,4] These standard protocols identify four relevant aspects of the evaluation: selecting materials to include in the evaluation, identifying the test procedures and associated analysis methods, developing the experimental matrix, and conducting a statistical analysis of the experimental results. Each of these aspects are described in the following sections. In addition to these activities, the research team also conducted supplementary experiments to further support efforts to refine the protocol. This chapter describes the details of the methodology followed in conducting the ruggedness and supplementary tests.

MATERIALS

Three laboratory-mixed, laboratory-compacted mixtures were used for the ruggedness study: a dense-graded 9.5-mm NMAS surface mixture from Alabama, a dense-graded 19-mm NMAS mixture from North Carolina, and a dense-graded 25-mm NMAS base mixture also from North Carolina. Details pertaining to the selected mixtures are provided in table 1, and the full job mix formula for each design is in the appendix. Figure 1 shows the gradations for selected mixtures.

These materials were chosen based on guidance in ASTM C1067-12, which states that the materials should be selected to cover the range of material properties for which the test method is applicable.[4] The range of material properties includes both the test measure itself and the variability of that measure. An unmodified mixture, a polymer-modified mixture, and an unmodified mixture with reclaimed asphalt pavement (RAP) were chosen to represent the scope of test measures and the range in NMAS values to represent the array of test variability. The procedures for AMPT cyclic fatigue testing using large specimens covers materials with NMAS of up to 25 mm, whereas the test procedure for small specimens is limited to mixtures with NMAS of up to 19 mm. Therefore, ruggedness evaluations for both the large and small specimens included the 9.5-mm NMAS mixture. However, the second mixture used in the ruggedness experiments on large and small specimens differed according to the maximum allowable NMAS. Polymer modification is typically used in surface mixtures; therefore, each ruggedness evaluation included an unmodified mixture with large NMAS and a polymer-modified mixture with small NMAS.

Table 1. Selected mixtures for the ruggedness evaluation.

Source	NMAS (mm)	Virgin Binder Grade	Polymer Modified?	Binder Content (%)	RAP Content (%)	Recycled Binder Ratio	VMA (%)	VFA (%)	# Design Gyrations
NC	25.0	PG 64-22	No	4.1	25	0.24	13.2	70.0	75
NC	19.0	PG 64-22	No	4.6	30	0.26	14.6	73.8	65
AL	9.5	PG 76-22	Yes	5.8	0	0	16.1	75.1	80
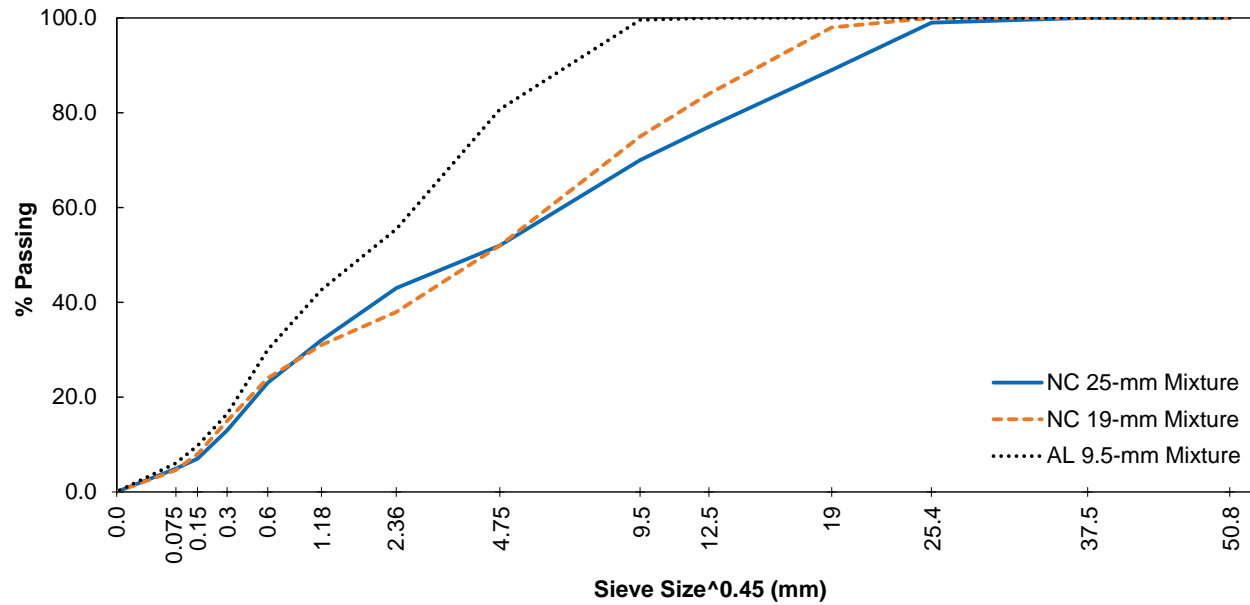
Source: FHWA.

Figure 1. Graph. Sieve size 0.45 mm power chart for gradations.

4

Two laboratories tested the 9.5-mm NMAS surface mixture, as suggested by ASTM C1067-12, while a single laboratory tested both the 19-mm NMAS and 25-mm NMAS mixtures. To best ensure uniformity, one operator fabricated all specimens for a given mixture type, and all specimens of a given type were randomized prior to distribution and testing.

TESTING AND ANALYSIS PROCEDURES

For this study, the research team carried out two separate experiments: $|E^*|$ tests and cyclic fatigue tests; however, the cyclic fatigue tests were the only tests subjected to the systematic change in factors. To avoid operator bias, a single operator conducted all experiments within a given laboratory and all data were analyzed at the lead researcher's facility.

Dynamic Modulus Testing

The research team conducted dynamic modulus tests of the large specimens following AASHTO R 83-17 and AASHTO T 378-17.[5,9] Dynamic modulus tests of the small specimens followed AASHTO PP 99-19 and AASHTO TP 132-19.[10,11] In both tests, asphalt specimens are subjected to frequency-sweep testing at 10, 1, and 0.1 Hz and at three temperatures. The large specimens were also subjected to 0.01 Hz testing at the highest test temperature. All specimens were tested at 4 and 20°C. The third temperature for the small geometry specimens and for the 25-mm NMAS specimens was 40°C. For the large 9.5-mm NMAS specimens, the third test temperature was 45°C because the third test temperature is determined by the binder performance grade (PG) in accordance with AASHTO R 84-17 (i.e., 40°C for PG 64-22 and 45°C for PG 76-22).[12]

The calculation of $|E^*|$, phase angle, and data-quality indicators followed the method outlined in each respective standard. However, there were some exceptions to the analysis of the modulus mastercurve data. AASHTO TP 107-18 and AASHTO TP 133-19 specify the use of the sigmoidal function to represent the storage modulus mastercurve of asphalt mixtures, as shown in figure 2. The lowest $|E^*|$ test temperature used for both small and large specimens is 4°C in accordance with AASHTO R 84-17 and AASHTO TP 132-19.[11,12] However, this temperature is not low enough to capture the glassy behavior of asphalt mixtures using the testing frequencies specified in the current standards. In addition, Keshavarzi has demonstrated that sigmoidal model coefficients optimized using AMPT $|E^*|$ test results lead to errors when used to extrapolate the predicted behavior at very high reduced frequencies approaching the glassy regime.[13] Olard and Di Benedetto as well as Keshavarzi have demonstrated that the coefficients of the two springs, two parabolic elements, one dashpot (2S2P1D) model, shown in figure 3, optimized using AMPT $|E^*|$ test results, can be used to accurately predict the behavior of asphalt mixtures at very high reduced frequencies approaching the glassy regime.[13,14] Correspondingly, the 2S2P1D model was used to represent the $|E^*|$ mastercurve herein.

$$\log(E'_{Sigmoidal}(\omega,T)) = \log(E'_{Sigmoidal}(\omega_R)) = b + \frac{\log(maxE') - b}{1 + e^{d + g \times \log(\omega_R)}}$$

Figure 2. Equation. Sigmoidal function for the storage modulus mastercurve.

Where:

$E'(\omega,T)$ = storage modulus at a particular temperature and angular frequency (kPa or psi).
$E'(\omega_R)$ = storage modulus at a particular reduced angular frequency (kPa or psi).
$E'_{Sigmoidal}$ = storage modulus predicted from the Sigmoidal model (kPa or psi).
ω = angular frequency (rad/s).
T = temperature (°C).
ω_R = reduced angular frequency (rad/s).
$max\ E'$ = maximum storage modulus.
b, d, g = fitting coefficients.

$$E'_{2S2P1D} = E_{00} + \frac{E'_1}{\left(\left(\frac{E'_1}{E_0 - E_{00}}\right)^2 + \left(\frac{E'_2}{E_0 - E_{00}}\right)^2\right)}$$

A. Storage modulus from 2S2P1D model.

$$E'_1 = (E_0 - E_{00}) \times \left(1 + \delta \times (\omega_R \times \tau_E)^{-\kappa} \times \cos\left(\frac{\kappa\pi}{2}\right) + (\omega_R \times \tau_E)^{-h} \times \cos\left(\frac{h\pi}{2}\right)\right)$$

B. Storage modulus subequation 1.

$$E'_2 = (E_0 - E_{00}) \times \left(\begin{array}{l}1 + \delta \times (\omega_R \times \tau_E)^{-\kappa} \times \sin\left(\frac{\kappa\pi}{2}\right) + (\omega_R \times \tau_E)^{-h} \times \sin\left(\frac{h\pi}{2}\right) + \\ (\omega_R \times \tau_E \times \beta)^{-1}\end{array}\right)$$

C. Storage modulus subequation 2.

Figure 3. Equation. 2S2P1D model.

Where:

E_0 = maximum storage modulus value (kPa or psi).
$\kappa,\ \delta,\ \gamma,\ h,\ \beta,\ \tau_E$ = fitting coefficients.
E'_{2S2P1D} = storage modulus from 2S2P1D model.

Analysis of the results of the cyclic fatigue testing also requires $|E^*|$ test results. To best reflect the effect of a difference in air-void content in cyclic fatigue testing, the $|E^*|$ test results at both air-void levels were analyzed together and the averaged result was coupled with all results of the cyclic fatigue testing obtained from each determination.

Cyclic Fatigue Testing

The research team conducted cyclic fatigue testing according to the experimental plan for the ruggedness evaluation. The team carried out all tests following either the AASHTO TP 107-18 (large specimen) or TP 133-19 (small specimen) protocol at 21°C. The test temperature was selected based on the guidance given in AASHTO TP 133-19, which specifies test temperatures based on the expected blended binder grade for the mixtures containing RAP. When a test resulted in a specimen failure outside the instrumented gauge points (i.e., end failure), the test

was considered invalid and an additional specimen was tested. The research team maintained a record of end failures and considered that record when judging the significance of individual evaluation factors.

Primary Analysis

The damage-characteristic curve constitutes the relationship between C, a material integrity indicator, and S. The research team followed the standard equations for calculating these quantities, as presented in the AASHTO TP 107-18 and TP 133-19 test protocols.[1,2] The team used FlexMAT[TM] 2.0 software, developed under FHWA sponsorship, for this purpose.[15] Some exceptions to other aspects of the equations in these protocols were followed. The first involves the analytical fitting function used to represent the damage-characteristic curve. According to AASHTO TP 107-18, the damage-characteristic curve should be reported using either the exponential or power law model given in figure 4 and figure 5, respectively.[1] AASHTO TP 133-19 specifies only the use of figure 5 based on the results of Lee et al., which demonstrated that the damage-characteristic curve is better represented by the power law model, as shown in figure 6.[2,16] In this ruggedness evaluation, the research team only used the power law model to analyze both small and large specimens because the results of Lee et al. demonstrate that it best represents the results of AMPT cyclic fatigue testing [16]

$$C = e^{a_1 S^{a_2}}$$

Figure 4. Equation. Exponential function of damage-characteristic curve.

In figure 4, a_1 and a_2 are fitting coefficients for the exponential model.

$$C = 1 - C_{11} S^{C_{12}}$$

Figure 5. Equation. Power law function of damage-characteristic curve.

In figure 5, C_{11} and C_{12} are fitting coefficients for the power law model.
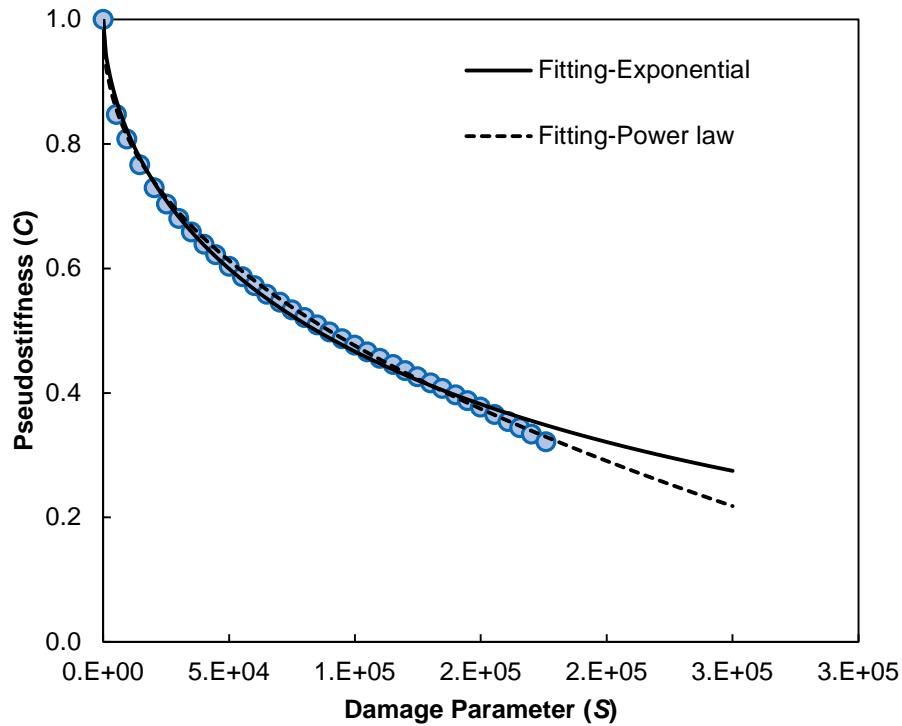
Source: FHWA.

Figure 6. Graph. Comparison of exponential and power law models for fitting the damage-characteristic curve.[16]

The failure criterion in AASHTO TP 107-18 is a power law function that relates the average rate of pseudostrain energy release (G^R) in a test to the number of cycles to failure (N_f). Because N_f cannot be directly controlled, individual specimen results cannot be regarded as replicates and thus cannot be evaluated statistically using the precise procedures given in ASTM E1169-14 or C1067-12.[3,4] In addition, because model parameters in the G^R failure criterion are obtained by linear regression in log–log scale, the deterministic fitted model coefficients are highly impacted by the test variability. Moreover, the G^R values, which are used in pavement-performance prediction, are beyond the range of laboratory testing, so extrapolation is necessary.[17] The extrapolation of a power function is highly sensitive to the power term, and the test outcome variability was further exaggerated by the extrapolation, as shown in figure 7.
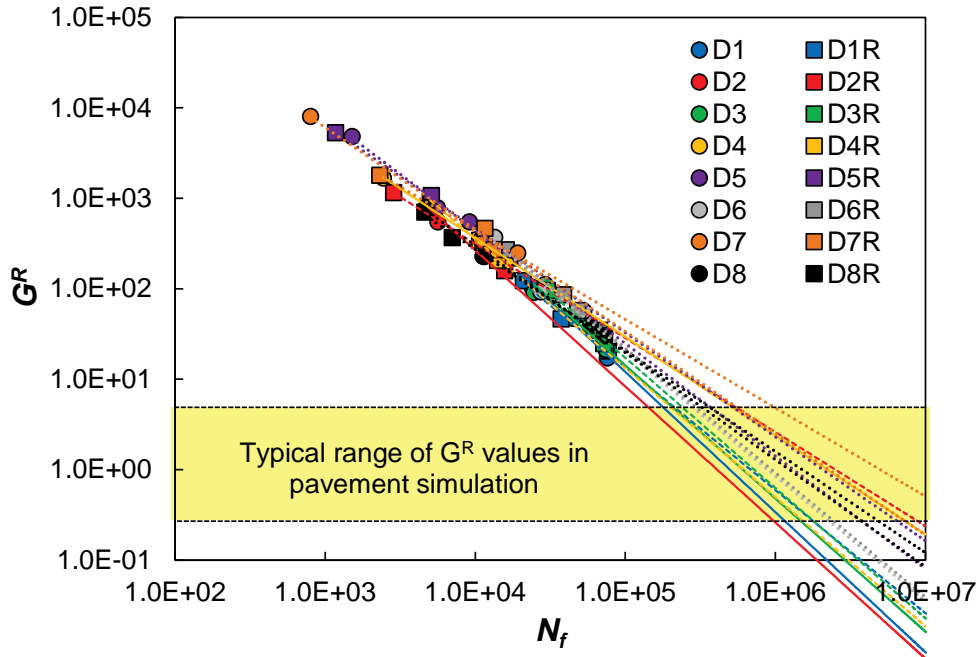
Source: FHWA.

Figure 7. Graph. G^R failure criteria for lab A 9.5-mm small-specimen testing.

In the ruggedness evaluation, this uncertainty can make a big impact on the statistical analyses at both material and pavement levels. The failure criterion used in the ruggedness evaluation, consistent with AASHTO TP 133-19 and defined in figure 8, is the average reduction in pseudostiffness up to failure.[2] The D^R of a given mixture is determined as the slope of the line formed by the relationship between the summation of $(1 - C)$ up to failure (i.e., $cum(1 - C)$) versus N_f, as shown in figure 9. Thus, the deviation of the individual data points from the best fit line is a measure of specimen-to-specimen variability in the failure criterion results.

$$D^R = \frac{\int_0^{N_f} (1 - C)dN}{N_f} = \frac{Cum(1 - C)}{N_f}$$

Figure 8. Equation. D^R failure criterion function.

In figure 8, N equals the number of cycles.

9
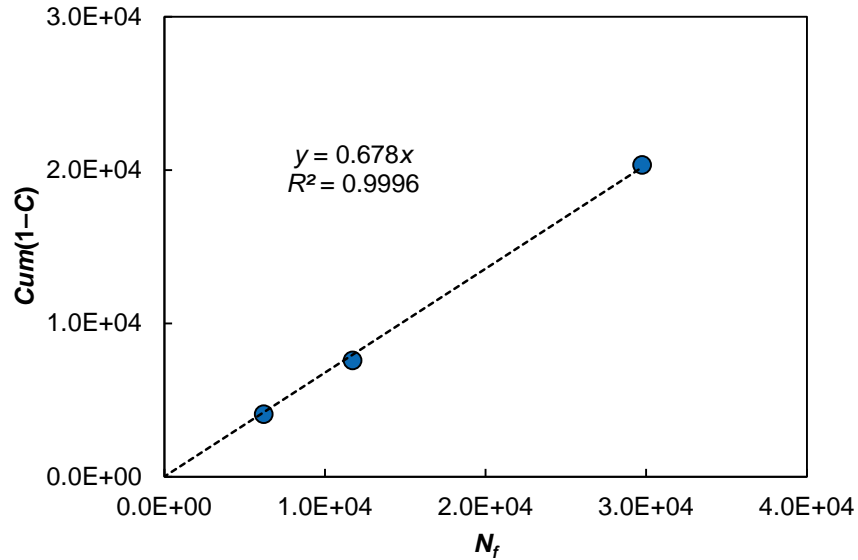
Source: FHWA.

Figure 9. Graph. D^R failure criterion.

The calculation of D^R requires defining N_f. AASHTO TP 107-18 and AASHTO TP 133-19 define N_f as the cycle in which the phase angle drops sharply after a stable increase during cyclic loading.[1,2] In most of the cyclic fatigue tests, a steady increase in phase angle regarding loading cycles is initially observed followed by a marked decrease, allowing for clear definition of the failure cycle. This decrease in phase angle results because, when localization happens, the linear variable differential transformer (LVDT) signal becomes distorted and the sinusoidal regression calculations used to define strain magnitude and phase angle poorly fit the measured data. However, as shown in figure 10, the phase angle may lack a well-defined peak if large discrepancies exist between the times when localization occurs at the different LVDT positions. During testing, three LVDTs are mounted on the specimen 120 degrees apart, and the aggregated LVDT results are used to report phase angle. When localization does not occur across the entire specimen at the same time, one LVDT may have lost its signal quality, but the others may maintain adequate quality such that the averaging process leads to a less-than-marked decrease in phase angle. Localization at even a single LVDT position should define specimen failure, and this smearing process represents a potential for error in determining the failure cycle.

10
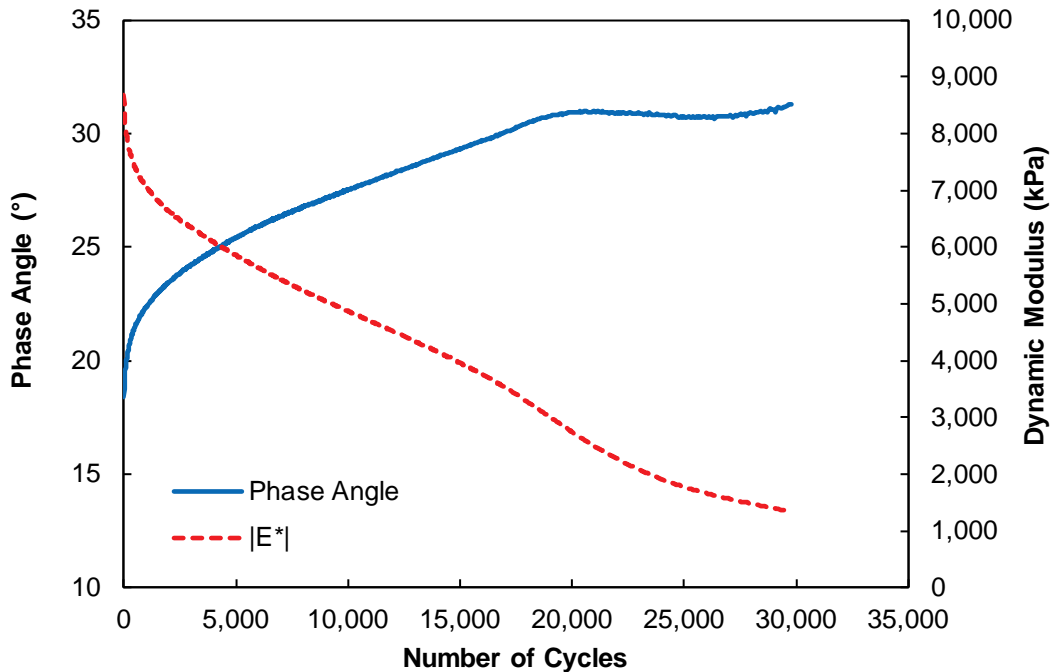
Source: FHWA.

Figure 10. Graph. Example of 19-mm small specimen ruggedness testing demonstrating a clear peak in Stress x *N*.

While this effect has been known since the earliest drafts of the AASHTO TP 107-18 protocol, it was neither widely seen nor systemic across mixtures and specimens. Likewise, the failure criterion was not overly sensitive to the errors in failure-cycle determination, and no clear alternative existed. However, the definition of N_f as the cycle where the product of stiffness and cycle number (*N*) reach a maximum value before declining, included in AASHTO T 321-17, provides a reasonable alternative.[18] Based on this change in AASHTO T 321-17, the research team proposed to use a refined failure definition for cyclic fatigue as the cycle where the product of peak-to-peak stress and cycle number reaches its highest value before declining. Figure 11 shows the pattern of the product of stress and cycle number versus cycle number for the same test shown in figure 10.

The use of stress amplitude instead of stiffness eliminates the need for any onspecimen LVDT measurements and allows the data analyst to define the failure cycle accurately even in cases where the specimen fails outside of the gauge points. In addition, this definition maintains some consistency with AASHTO T 321-17 because in a test where the strain level is constant, the point where the product of the stiffness and cycle number is maximized is the same as the point where the product of stress amplitude and cycle number is maximized. The research team analyzed the effect of changing the failure definition by applying both the defined peak in phase angle and the product of stress and defined cycle number to all the ruggedness mixtures with both large- and small-specimen geometries and NMAS values ranging from 9.5- to 25-mm. The results from this analysis are summarized in figure 12, which shows that the peak of stress multiplied by the number of cycles yields a reduction in fatigue life of approximately 6 percent

in the ruggedness-experiment results, as the slope of the regression line (0.9415) indicates. Since the new failure definition results in a failure cycle less than that of the peak in the defined phase angle, the data from the tests with the old failure criterion contain all the data before the failure according to the new failure criterion. Therefore, previous data can be easily reanalyzed and reinterpreted with the new definition.
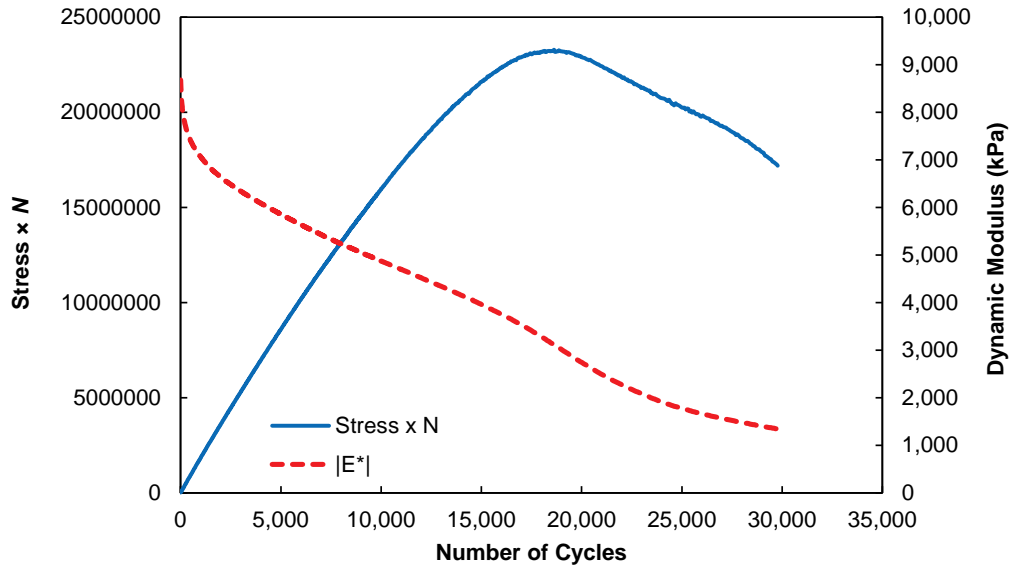
Source: FHWA.

Figure 11. Graph. Example of ruggedness testing on a 19-mm small specimen showing unclear failure cycle based on the peak in phase angle.
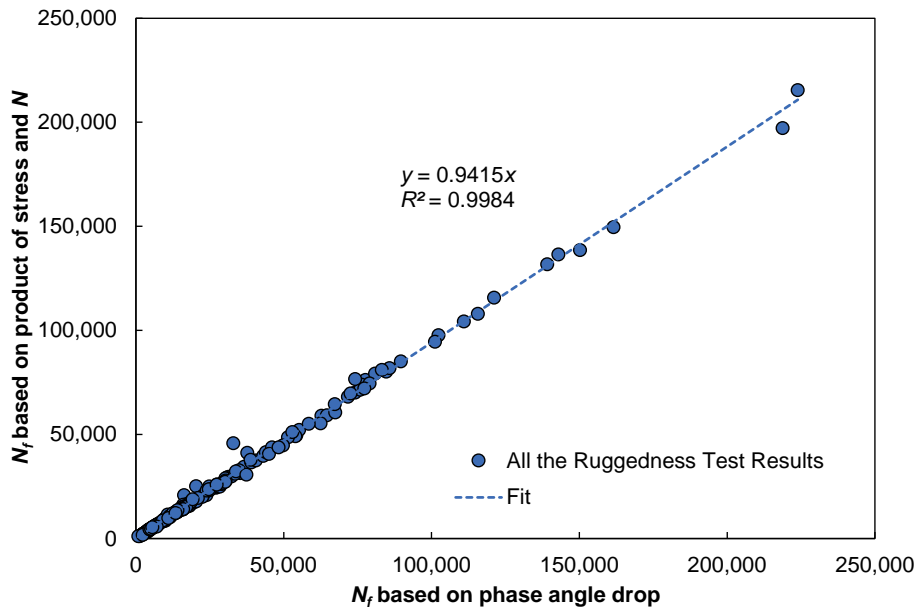
Source: FHWA.

Figure 12. Graph. N_f based on peak in phase angle and product of stress and N.

12

Apparent Damage Capacity

An index parameter to indicate asphalt mixture fatigue resistance, termed the apparent damage capacity (S_{app}) and defined in figure 13, can be calculated using AMPT cyclic fatigue test results.[19] The S_{app} value accounts for the effects of material's modulus and toughness on its fatigue resistance and is a measure of the amount of fatigue damage the material can tolerate under loading. The temperature for the S_{app} calculation is equal to the average of high- and low-temperature grades at the standard 98-percent reliability level minus 3°C for the climate where the mixture is to be placed. Higher S_{app} values indicate increased fatigue resistance. The S_{app} value is sensitive to mixture factors (e.g., aggregate gradation, asphalt binder content, RAP content, and asphalt binder grade), compaction, and aging and meets general expectations regarding the effects of these parameters on fatigue-cracking performance.[19] While S_{app} is not included in the current standards for cyclic fatigue testing, it is considered a key test outcome and thus was evaluated.

$$S_{app} = 1000^{\frac{\alpha}{2}-1} \frac{a_T^{\frac{1}{\alpha+1}} \left(\dfrac{D^R}{C_{11}} \right)^{\frac{1}{C_{12}}}}{\left| E^* \right|^{\frac{\alpha}{4}}}$$

Figure 13. Equation. S_{app} definition.

Where:
   α = damage growth rate.
   a_T = time–temperature shift factor at a given temperature.

Predictions of Pavement Performance

To evaluate the combined effects of damage-characteristic curve and failure criterion results, the research team conducted pavement performance predictions using FlexPAVE™ version 1.1.[20] FlexPAVE, developed under the sponsorship of FHWA, is a pavement-performance analysis software package that uses the S-VECD model to integrate the effects of loading rate and temperature into a structural model that then computes the pavement's responses and damage evolution. The FlexPAVE program calculates the long-term fatigue damage and rut depth of asphalt pavements under changing environmental and loading conditions. The output from FlexPAVE simulations of fatigue cracking is the percentage of damage, which is computed as the ratio of the damaged cross-section area to the total effective cross-sectional area. The total effective area is defined by using two overlapping triangles. The upper triangle is inverted and has a 170-cm-wide base located at the top of the surface asphalt layer and a vertex located at the base of the bottom asphalt layer. The 119-cm-wide base of the second (upright) triangle is located at the base of the bottom asphalt layer and its vertex is positioned at the top of the surface asphalt layer. The shape and dimensions of the effective area were established by examining damage contours for multiple simulations encompassing different mixtures and pavement structures where it was observed that damage never occurred outside of this area.[13,21]

In this study, a pavement structure consisting of a 10-cm layer of asphalt mixture over a 20-cm aggregate base and 380-cm subgrade was used to evaluate the results of the surface mixture. The research team carried out the simulation using enhanced integrated climatic model data for Raleigh, NC. The traffic level was set to 1,400 daily equivalent-single-axle loads (ESALs) with no growth. Another simple pavement structure, consisting of a 5-cm asphalt mixture surface layer over a 10-cm asphalt mixture intermediate or base layer, 20-cm aggregate base, and 380-cm subgrade was used for the analysis of the intermediate/base mixtures. The traffic level was set to 4,200 daily ESALs. The research team chose these traffic levels so that the resultant percentage of cracking (after application of the transfer function) would be within the range of 15 to 25 percent, which constitutes the region most sensitive to changes in material properties and therefore is most sensitive to the changes in the ruggedness factors.[21]

While the percentage of damage predicted by FlexPAVE accounts for the combined effects on pavement structural level analysis, there is, however, a limitation on how much it can represent the visual assessment of performance in the field. Fatigue performance in the field is measured by visible cracking on the pavement surface. However, the percentage of damage determined in FlexPAVE is calculated based on the pavement cross section. To overcome this limitation in the ruggedness study, a transfer function, which was established based on 39 pavement sections from 4 field projects (the details of which are summarized in table 2), was used to convert the percentage of damage in the cross section of the asphalt layer (*%Damage*) to the percentage of cracking (*%Cracking*). The *%Cracking* is the ratio of the sum of the fatigue-cracking area to the total lane area, where the fatigue-cracking area is the sum of the fatigue-cracking area and the area affected by longitudinal cracking in the wheel paths.[21] The area affected by longitudinal cracking is calculated as the product of the length of the crack and the affected range, which is 0.3 m. Figure 14 provides the transfer function equation used in this ruggedness study, and figure 15 shows the relationship between the measured percentage of surface cracking observed from field sections and the *%Cracking* predicted by FlexPAVE simulations.[21]

Table 2. Field project used to establish a transfer function.

Field Project	Number of Sections	Location
KEC Test Road	16	South Korea
NCAT Test Track	5	Alabama, US
MIT-RAP Test Road	4	Manitoba, Canada
MnROAD Test Road	14	Minnesota, US

KEC = Korea Expressway Corporation; MIT-RAP = Manitoba Infrastructure and Transportation Reclaimed Asphalt Pavement; MnROAD = Minnesota roadway test track; NCAT = National Center for Asphalt Technology.

$$\%Cracking = \frac{50}{1 + \dfrac{7{,}970{,}327}{\%Damage^{6.0672}}}$$

Figure 14. Equation. Fatigue transfer function.
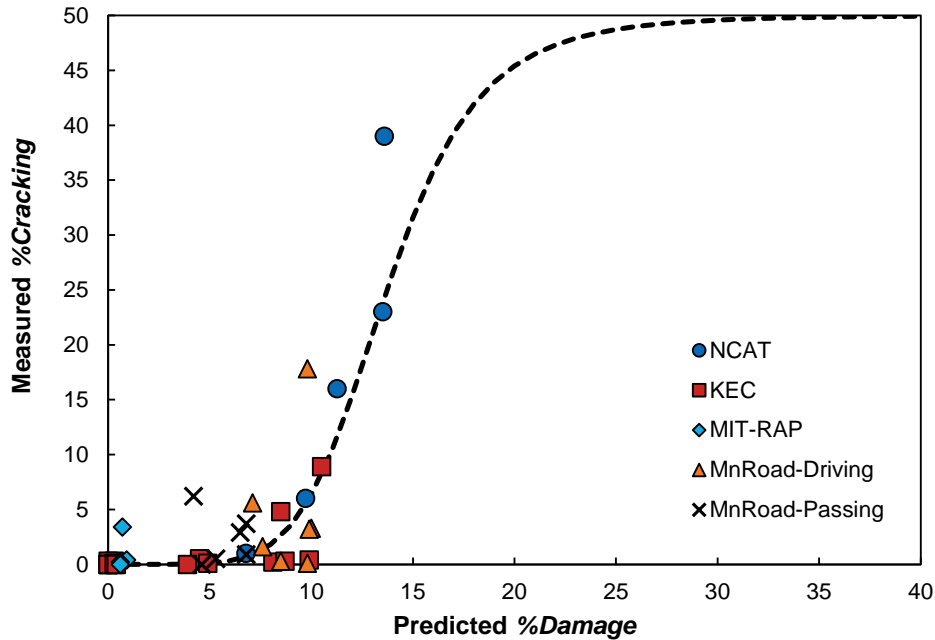
Source: FHWA.

Figure 15. Graph. Surface measured *%Cracking* versus predicted *%Damage*.

EXPERIMENTAL MATRIX

Per ASTM E1169-14 and C1067-12, the ruggedness study was carried out according to the Plackett–Burman experimental design.[3,4,22] This designed experiment allows factors to be evaluated at two levels. These levels are selected to reflect the ranges for each factor that would reasonably be selected given imprecise guidance in the standard. The research team included seven factors in the ruggedness evaluation of the AMPT cyclic fatigue tests, which allowed them to accomplish the ruggedness study using 16 determinations, including replication. Replication is suggested by ASTM E1169-14 and required by ASTM C1067-12 to improve precision when assessing the statistical significance of evaluation factors.[3,4] Each determination included three cyclic fatigue tests per AASHTO TP 107-18 and AASHTO TP 133-19.[1,2]

Table 3 summarizes the general layout of the seven-factor design with each factor's corresponding levels. Table 4 through table 7 provide the specific conditions used for each mixture, specimen geometry, and determination. Justifications for these experimental factors and the basis for selecting the levels follow. The research team used guidance in ASTM E1169-14 and ASTM C1067-12 to select the levels for each of these factors. In addition, when preparing test specimens, the team controlled the levels for each factor to best mimic the procedures a test operator would follow given the AMPT specimen preparation equipment.

15

Table 3. Ruggedness experimental plan.

Factor	Determination*							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Air voids	a	a	a	a	A	A	A	A
Specimen height	b	b	B	B	b	b	B	B
Loading-platen parallelism	C	c	C	c	C	c	C	c
Difference in diameter of loading platen and specimen (mm)	D	D	d	d	d	d	D	D
Strain level	e	E	e	E	E	e	E	e
Rest period	F	f	f	F	F	f	f	F
Temperature	g	G	G	g	G	g	g	G

*Capital letters indicate level 1 factors and lowercase letters indicate level 2 factors.

Table 4. Experimental plan for large-specimen testing of the 9.5-mm NMAS mixture.

Factor	Determination							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Air voids (%)	7.0 ± 0.5	7.0 ± 0.5	7.0 ± 0.5	7.0 ± 0.5	5.0 ± 0.5	5.0 ± 0.5	5.0 ± 0.5	5.0 ± 0.5
Specimen height (mm)	135	135	125	125	135	135	125	125
Loading-platen parallelism (degree)	0	0.6 (T) 0.3 (S)	0	0.6 (T) 0.3 (S)	0	0.6 (T) 0.3 (S)	0	0.6 (T) 0.3 (S)
Difference in diameter of loading platen and specimen (mm)	−4	−4	+7	+7	+7	+7	−4	−4
Strain level (microstrain)	370, 330, 300 (T); 430, 380, 330 (S)	470, 420, 370 (T); 540, 480, 430 (S)	370, 330, 300 (T); 430, 380, 330 (S)	470, 420, 370 (T); 540, 480, 430 (S)	470, 420, 370 (T); 540, 480, 430 (S)	370, 330, 300 (T); 430, 380, 330 (S)	470, 420, 370 (T); 540, 480, 430 (S)	370, 330, 300 (T); 430, 380, 330 (S)
Rest period (min)	5	20	20	5	5	20	20	5
Temperature (°C)	19	23	23	19	23	19	19	23

(T) = level used for lab B tests; (S) = level used for lab A tests.

Table 5. Experimental plan for large-specimen testing of the 25-mm NMAS mixture.

Factor	Determination							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Air voids (%)	6.0 ± 0.5	6.0 ± 0.5	6.0 ± 0.5	6.0 ± 0.5	4.5 ± 0.5	4.5 ± 0.5	4.5 ± 0.5	4.5 ± 0.5
Specimen height (mm)	135	135	125	125	135	135	125	125
Loading-platen parallelism (degree)	0	0.3	0	0.3	0	0.3	0	0.3
Difference in diameter of loading platen and specimen (mm)	−4	−4	+7	+7	+7	+7	−4	−4
Strain level (microstrain)	190, 175, 165	220, 205, 190	190, 175, 165	220, 205, 190	220, 205, 190	190, 175, 165	220, 205, 190	190, 175, 165
Rest period (min)	5	20	20	5	5	20	20	5
Temperature (°C)	19	23	23	19	23	19	19	23

Table 6. Experimental plan for small-specimen testing of the 9.5-mm NMAS mixture.

Factor	Determination							
	1	2	3	4	5	6	7	8
Air voids (%)	7.0 ± 0.5	7.0 ± 0.5	7.0 ± 0.5	7.0 ± 0.5	5.0 ± 0.5	5.0 ± 0.5	5.0 ± 0.5	5.0 ± 0.5
Specimen height (mm)	112.5	112.5	107.5	107.5	112.5	112.5	107.5	107.5
Loading-platen parallelism (degree)	0	0.6	0	0.6	0	0.6	0	0.6
Difference in diameter of loading platen and specimen (mm)	−2 mm	−2 mm	+3 mm	+3 mm	+3 mm	+3 mm	−2 mm	−2 mm
Strain level (microstrain)	590, 500, 430 (T); 660, 580, 510 (S)	810, 690, 590 (T); 840, 740 660 (S)	590, 500, 430 (T); 660, 580, 510 (S)	810, 690, 590 (T); 840, 740 660 (S)	810, 690, 590 (T); 840, 740 660 (S)	590, 500, 430 (T); 660, 580, 510 (S)	810, 690, 590 (T); 840, 740 660 (S)	590, 500, 430 (T); 660, 580, 510 (S)
Rest period (min)	5	20	20	5	5	20	20	5
Temperature (°C)	19	23	23	19	23	19	19	23

(T) = level used for lab B tests; (S) = level used for lab A tests.

Table 7. Experimental plan for small-specimen testing of 19-mm NMAS mixture.

Factor	Determination							
	1	2	3	4	5	6	7	8
Air voids (%)	5.5 ± 0.5	5.5 ± 0.5	5.5 ± 0.5	5.5 ± 0.5	4.0 ± 0.5	4.0 ± 0.5	4.0 ± 0.5	4.0 ± 0.5
Spec. Height (mm)	112.5	112.5	107.5	107.5	112.5	112.5	107.5	107.5
Loading-platen parallelism (degree)	0	0.6	0	0.6	0	0.6	0	0.6
Difference in diameter of loading platen and specimen (mm)	−2	−2	+3	+3	+3	+3	−2	−2
Strain level (microstrain)	230, 215, 200	265, 250, 230	230, 215, 200	265, 250, 230	265, 250, 230	230, 215, 200	265, 250, 230	230, 215, 200
Rest period (min)	5	20	20	5	5	20	20	5
Temperature (°C)	19	23	23	19	23	19	19	23

Air-Void Content

In a testing procedure, tolerance controls for air voids prescribe the allowable range within which the procedure produces a consistent test result. The range is ideally as large as possible to limit the rejection of test specimens. The ruggedness study conducted for AASHTO T 378-17 did not include air-void content as a level for two reasons: first, previous data suggested that the errors in $/E*/$ tests and flow number tests that stem from a 1-percent change in air-void content could be significant, and second, control tighter than a ±0.5-percent tolerance is infeasible given the variability in specimen fabrication procedures, according to Bonaquist.[5] The limited data from the NCHRP IDEA N-181 project *Development of Small Specimen Geometry for Asphalt Mixture Performance Testing* suggested that the test results from specimens outside of the ±0.5-percent tolerance may be acceptable.[23] If this finding were then verified, it would be advantageous because a wider tolerance range for air voids would produce fewer rejected specimens and deliver a rugged testing standard. Initially, the research team intended to use a target air-void content of 7 ± 0.5 percent as the high level and 5 ± 0.5 percent for the low level for each mixture.

However, the team encountered a high occurrence of end failures at an air-void content of 7 ± 0.5 percent in 19- and 25-mm mixtures. At the 7 ± 0.5 percent level, the larger NMAS mixtures required fewer than 10 gyrations to compact 180-mm-tall gyratory-compacted samples. The team speculated that a low compaction effort may exacerbate the air-void gradient in gyratory-compacted samples. Therefore, the research team selected lower air-void-content levels for 19- and 25-mm mixtures to ensure the number of gyrations required to compact the 180-mm-tall sample exceeded 10. The maximum level of air-void content to ensure greater than 10 gyrations of compaction effort corresponded to 6 percent for the 25-mm mixture and 5.5 percent for the 19-mm mixture. The low air-void levels for these mixtures were selected as 4.5 percent and 4 percent for the 25-mm and 19-mm mixtures, respectively.

Specimen Height

AASHTO TP 107-18 specifies a specimen height of 130 ± 2.5 mm based on experimental and finite element modeling analyses.[24] The height tolerance in AASHTO TP 107-18 (± 2.5 mm) was adopted based on the height tolerance for fabricating 150-mm-tall specimens in AASHTO R 83-17 but without experimental basis.[9] However, prior tests indicate that taller specimens may have higher air-void contents at the top and bottom and thus experience failure at the specimen ends located outside the instrumented gauge points.[24] By contrast, small specimens are extracted from the middle 110 mm of 180-mm tall gyratory samples and so are not expected to exhibit high air-void gradients. However, taller specimens have been found to yield different results, and shorter specimens may not fit inside the AMPT using standard fixtures.[24]

Loading-Platen Parallelism

Tensile strains are induced in a test specimen if the loading platens are not parallel with the AMPT platens when the attachment bolts are tightened. Loading-platen parallelism issues can result from improper cleaning or damage to the gluing jig due to either repeated use of the jig or poor gluing practices. It can also occur because of misalignment within the AMPT due to use (i.e., wear of the seals and other mechanical elements in the actuator) or within manufacturing tolerance differences. AASHTO TP 107-18 and TP 133-19 address this issue by requiring the use of a shim or ball bearing if a gap between the loading platen and top support of the AMPT exceeds 1 mm. However, this 1-mm limit lacks theoretical or experimental basis. A 1-mm gap between the loading platen and the top support of the AMPT at one side of the specimen would cause a tensile strain of up to 5,917 $\mu\varepsilon$ (i.e., microstrains) on one side of the 100-mm diameter specimen when the attachment bolts are tightened. The lower level of this factor was set at 0 degrees, consistent with the manufacturer's guidance on usage of the loading platens and gluing jig. Initially, the second level of this factor was intended to be 0.6 degrees for both large- and small-specimen testing. However, the initial large specimen ruggedness evaluation conducted by lab B using the 9.5-mm NMAS mixture demonstrated a high occurrence of end failure when using the 0.6-degree platen angle. Consequently, the loading-platen angle was lowered to 0.3 degrees in subsequent ruggedness evaluations using large specimens conducted by lab A. The 0.6-degree level was used for all small-specimen testing, however.

The AMPT gluing jig was modified at the research facility's precision machine shop to control loading-platen parallelism in the ruggedness study.

Differences in Diameter of the Loading Platen and Specimen

Differences between the diameters of the loading platen and specimen may affect the operator's ability to center the specimen when attaching the loading platen, which in turn can impact load transfer from the platen to the specimen during testing. Given the combination of specimen fabrication standards (i.e., AASHTO PP 99-19 and AASHTO R 83-17) and the standards for cyclic fatigue testing, a relatively wide range of loading-platen diameters is currently permitted, including where the loading-platen diameter is larger than the specimen diameter and vice versa.[9,10] This factor can be managed by using core bits to control the specimen diameter, machining the loading platens, or both.

Strain Level

The standards for AMPT cyclic fatigue testing require three replicate tests per mixture, all conducted at the same temperature and frequency, to characterize fatigue performance. The testing standards give guidance for the target N_f of the test specimens. AASHTO TP 133-19 specifies the rejection of test results where N_f is not within the range of 2,000 to 80,000 cycles. The strain levels indicated in the experimental matrices in table 4 through table 7 were selected using pilot testing (described in the Supplementary Experiments section) to determine the input strain levels that would yield approximately 5,000; 10,000; and 20,000 cycles to failure for level 1 and 20,000; 40,000; and 80,000 cycles for level 2.

Rest Period

AASHTO TP 107-18 and AASHTO TP 133-19 require a 20- to 45-min rest period between the end of the fingerprint test and start of the cyclic fatigue test.[1,2] The rest period was included in the original AASHTO TP 107-18 testing procedure based on older servohydraulic testing systems where the actuator can move slightly during the transition from load- to displacement-control mode. The inclusion of a rest period was intended to allow time for the induced stresses to relax prior to starting the cyclic fatigue test. However, monitoring the load in the AMPT during the transition from load- to displacement-control mode demonstrated the absence of any loading to the specimen. Therefore, the rest period is only necessary to allow for recovery following the fingerprint test. Dynamic modulus testing standards suggest that the time required for recovery between the fingerprint and cyclic fatigue testing may be shorter than the period reflected in AASHTO TP 107-18 and TP 133-19. AASHTO T 378-17 does not require rest periods between $|E^*|$ testing frequencies, but AASHTO T 342-11 specifies a maximum 30-min rest-period duration between loading frequencies and states that a typical rest-period duration is 2 min.[6,25] The current default procedures for AMPT cyclic fatigue testing maintains zero load throughout the rest period and transitions to displacement-control mode just before beginning the cyclic fatigue testing. Given the ability of the AMPT to transition between modes of loading without inducing loading to the specimen, the ruggedness-evaluation experiments maintained zero load during the rest period and evaluated the effect of rest-period duration.

Temperature

Errors are introduced when using an incorrect temperature to analyze cyclic fatigue tests. Incorrect temperatures may be recorded if the specimen has not fully equilibrated, which can

occur due to errors in conditioning and transfer procedures. Reproducing these errors in a ruggedness study in a broadly applicable way is not possible. However, it is possible to simulate the effects of such errors by artificially introducing different temperatures in the damage curve calculations. To simulate the effect, the research team conducted ruggedness experiments at 21°C. The team then artificially modified measured temperatures in the output data based on the ruggedness experimental plans and used FlexMAT to analyze the test data.

The research team altered the calculation of the Dynamic Modulus Ratio (*DMR*), defined in figure 16, according to the modified output files. The dynamic modulus from the 2S2P1D model ($|E^*|_{LVE}$) in this equation is calculated according to the temperature in the modified output file. Reduced time, defined in figure 17 and used in the calculation of damage, is also a function of the temperature included in the output file. The time–temperature shift factor in this equation is a function of temperature, so changing the temperature will affect the calculation of reduced time. The altered *DMR* and reduced time are also used in further calculations, such as pseudosecant modulus and damage parameter, as shown in figure 18 through figure 21. These factors affect the damage-characteristic curve.

$$DMR = \frac{|E^*|_{fingerprint}}{|E^*|_{LVE}}$$

Figure 16. Equation. *DMR* calculation.

In figure 16, $|E^*|_{fingerprint}$ is defined as the dynamic modulus determined from fingerprint testing.

$$t_R = \frac{t}{a_T}$$

Figure 17. Equation. Reduced-time calculation.

Where:
  t = time measured from the experiment (s).
  t_R = reduced time (s).

$$C = \frac{\sigma}{\varepsilon^R \times DMR}$$

Figure 18. Equation. Pseudosecant modulus calculation for initial half cycle.

Where:
  σ = stress (kPa or psi).
  ε^R = pseudostrain.

$$\Delta S_i = \begin{cases} \left(-\dfrac{DMR}{2} \left(\varepsilon^R \right)^2 \left(C_i - C_{i-1} \right) \right)^{\alpha/_{\alpha+1}} \left(\Delta t_R \right)^{1/_{\alpha+1}} & C_i \leq C_{i-1} \\ 0 & C_i > C_{i-1} \end{cases}$$

Figure 19. Equation. Damage-parameter calculation for initial half cycle.

Where:
ΔS_i = damage growth between the current and previous time step.
C_i = pseudosecant modulus at the current time step.
C_{i-1} = pseudosecant modulus at the previous time step.

$$C^* = \frac{\sigma_{pp}}{\varepsilon^R_{pp} \times DMR}$$

Figure 20. Equation. Cyclic pseudosecant modulus calculation for each cycle.

Where:
C^* = cyclic pseudosecant modulus.
σ_{pp} = peak-to-peak stress (kPa or psi).
ε^R_{pp} = peak-to-peak pseudostrain.

$$\Delta S_n = \begin{cases} \left(-\dfrac{DMR}{2} \left(\varepsilon^R_{ta} \right)^2 \left(C^*_n - C^*_{n-1} \right) \right)^{\alpha/_{\alpha+1}} \left(\Delta t_R \right)^{1/_{\alpha+1}} \left(K_1 \right)^{1/_{\alpha+1}} & C^*_n \leq C^*_{n-1} \\ 0 & C^*_n > C^*_{n-1} \end{cases}$$

Figure 21. Equation. Damage-parameter calculation for each cycle.

Where:
ΔS_n = damage growth between the current and previous analysis cycle.
ε^R_{ta} = tension amplitude of the pseudostrain.
K_1 = form adjustment factor.
C^*_n = cyclic pseudosecant modulus at the current analysis cycle.
C^*_{n-1} = cyclic pseudosecant modulus at the previous analysis cycle.

These temperature-related errors can be evaluated with two different methods:

1. Every specimen is tested at a single temperature, 21°C, and analyzed at either 19°C or 23°C, depending on the determination of interest (recall that the evaluation levels for the temperature factor are ±2°C); or
2. Depending on the determination of interest, specimens are tested at 19°C or 23°C but are analyzed at a single test temperature, 21°C.

For the first scenario, the effect of temperature error is introduced from the difference between 21°C (test temperature) and 19°C or 23°C (analysis temperature). In contrast, in the second method, the effect is introduced from the difference between 19°C and 23°C (test temperature) and 21°C (analysis temperature). Thus, while the actual temperature differences are similar

(±2°C), the effects may be slightly different (since, for example, the effect of a lower temperature in scenario one is assessed by the difference between 21°C and 19°C and in the second scenario it is the difference between 23°C and 21°C). This subtle distinction means that the effect of *DMR* will be slightly different for the two methods. Similarly, small differences in the magnitude of the time–temperature shift factor from 19°C to 21°C and then from 21°C to 23°C also exist. However, when the analysis temperature is lower than the test temperature, the effect from one method will be the same as when the analysis temperature is higher than the test temperature in the other method, but in the opposite direction. In addition, the statistical analysis in the ruggedness study uses the effect of the relative values from two different evaluation levels with linear relationship because the Plackett–Burman experimental design has two levels for each factor.[3] Therefore, the statistical analysis results from both methods will be the same. In this ruggedness study, the research team used the first method because the linearity of the effects can be figured out by artificially introducing different temperatures (e.g., ±0.5°C, ±4°C,) for further analysis.

STATISTICAL ANALYSIS

The AASHTO TP 107-18 and TP 133-19 test outcomes include the damage-characteristic function, and the research team performed the D^R-based failure criterion and statistical analysis on both. The statistical analysis included the following:

- Half-normal plots and Student's *t*-tests, which together assess the significance of factors based on the difference found in the average test results at the high and low levels as per ASTM E1169-14.[3]
- *F*-tests, which assess the difference in the variance of the test results between the high and low levels as per ASTM C1067-12.[4]

Both sets of statistical measures were used because each has precedence in ruggedness-testing studies on asphalt materials. The ruggedness protocols prescribe analyses for single measurements (e.g., cycles to failure). However, it is critical to note that the results of using the procedures for AMPT cyclic fatigue testing pose unique challenges regarding applying standard ruggedness protocols because one of the test results is a functional relationship (i.e., the damage-characteristic curve). To evaluate the ruggedness test results as directly as possible, two metrics were used to evaluate the damage-characteristic curve results: the values of *C* at three specific *S* levels and the area above the damage-characteristic curve. Based on those two metrics, the damage-characteristic curves, which are functional relationships, can be represented as single measurements and then used in the standard statistical analyses. The research team selected three specific *S* values to correspond to low-, intermediate-, and high-damage values in the test results. The high *S* level was selected based on the determination with the shortest damage-characteristic curve and hence, lowest maximum *S* value. The intermediate *S* value was selected to be close to the average of the low and high *S* values. The D^R failure criterion yields a single-point measurement; hence, the statistical analysis was applied directly to the test results. Finally, the research team evaluated the statistical significance of the factors on the apparent damage capacity, S_{app}. This index parameter was included because, although it is not a test outcome of the test protocols, it is likely to be a key application of the test outcomes. In addition, it is a single point value and thus is amenable to the statistical analysis in ASTM E1169-14 and ASTM C1067-12.[3,4]

THRESHOLD ANALYSIS

The statistical analysis identifies the factors that are statistically significant, but it does not inform the limit or threshold values to place on these factors. To define these limits, the research team used FlexPAVE because the results can be translated into percentage of cracking in pavement, an intuitive and practically important test outcome. Figure 22 shows the overall steps from mixture testing to the calculation for *%Cracking*. An acceptance range of ±10-percent error in the predicted *%Cracking* was used as the basis for defining the tolerance ranges for statistically significant experimental factors (e.g., if the *%Cracking* for the ideal condition is 20 percent, the range will be 18 to 22 percent). The methodology followed to use the FlexPAVE simulations involved multivariate linear regression. ASTM C1067-12 recommends using this type of approach to quantify the relative effect of changes in the level of factors when several are found to be statistically significant based on *F*-test analysis. It is a secondary benefit that the functions can also be used to establish threshold limits. This approach was used for the ruggedness evaluation in AASHTO T 378-17 and was adopted here because the results of this research show that multiple factors produce statistically significant effects.[5] The linear functions take the form shown in figure 23.
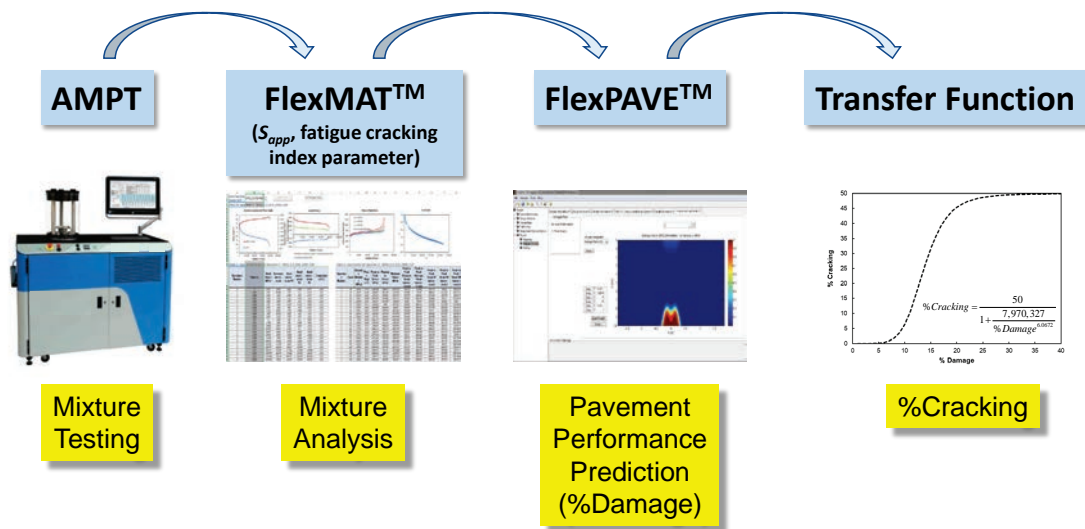
Source: FHWA.

Figure 22. Diagram. Overall steps from mixture testing to calculate *%Cracking*.

$$Y = B_0 + B_1X_1 + B_2X_2 + \cdots + B_7X_7 + Error$$

Figure 23. Equation. Multivariable linear regression equation.

Where:
    Y = *%Cracking* from FlexPAVE.
    X_{1-7} = seven factors in the ruggedness testing.
    B_{0-7} = threshold regression model coefficients.
    $Error$ = threshold regression model error.

Since each factor has different dimensionality and magnitude, the regression analysis is performed by normalizing the X_i values to the range of −1 to 1, where −1 and 1 represent the

lower and upper limit of the factor levels from the ruggedness experiments, respectively. As described in the next chapter, the research team performed this regression analysis on a determination-by-determination basis and on a specimen-by-specimen basis.

To establish the tolerance ranges for the significant factors, the research team followed a five-step process:

1. Perform the regression analysis by using all ruggedness-evaluation factors and, based on the resultant *p* values, determine the significant factors.
2. Rerun the regression analysis by using only significant factors identified from step 1 and establish a new regression model.
3. Based on the regression model from step 2, predict the *%Cracking* for an ideal condition in terms of the experimental factors. Define the ideal condition using the regression model by assigning a zero value for each factor except in the case of loading-platen parallelism, because −1 was assigned for the parallel level, which is the ideal condition.
4. Calculate two different values for *%Cracking* by using ±10-percent error from the *%Cracking* in the ideal condition (e.g., if the *%Cracking* for the ideal condition is 20 percent, the range will be 18–22 percent).
5. Based on the ±10 percent error limits in *%Cracking*, determine the allowable variations among individual factors using the regression model established in step 2. Translate the allowable variations, which were scaled based on ruggedness-evaluation levels, to the tolerance ranges for each factor.

The research team applied this regression approach for each ruggedness mixture and geometry and used it to propose tolerance ranges for the experimental factors in the specification.

SUPPLEMENTARY EXPERIMENTS AND ANALYSES

Proficiency Experiments

Prior to ruggedness testing, proficiency $|E^*|$ and cyclic fatigue experiments were carried out to ensure uniformity between laboratories A and B. These proficiency experiments were conducted using the 9.5-mm NMAS mixture acquired for the ruggedness evaluation. The air-void content for the tests was 7 ±0.5 percent, and tests were performed using both the small- and large-specimen geometries. For the large-specimen geometry, three $|E^*|$ tests were performed in each laboratory according to AASHTO T 378-17,[6] and four cyclic fatigue tests were conducted in each laboratory using AASHTO TP 107-18.[1] For the small-specimen geometry, three $|E^*|$ tests were performed in each laboratory according to AASHTO TP 132-19,[11] and four cyclic fatigue tests were performed in each laboratory according to AASHTO TP 133-19.[2] The results of the proficiency experiments are presented in the appendix.

Pilot Experiments

The labs conducted pilot cyclic fatigue tests for each mixture and geometry combination to establish strain levels for the ruggedness evaluation. For these pilot experiments, specimens were prepared at the higher air-void level. In addition, the labs performed cyclic fatigue tests at four different strain levels and relationships between the input strain levels, establishing the resultant N_f. Using these relationships, the ruggedness strain levels were selected to yield N_f values of

approximately 5,000; 10,000; and 20,000 for level 1 factors and 20,000; 40,000; and 80,000 for level 2 factors. The results of the pilot experiments are presented in the appendix.

End-Failure Mitigation Study

The experiments revealed that a lack of parallelism in the loading platen led to a high occurrence of end failure in cyclic fatigue tests. Therefore, the research team determined that a process to allow the attachment of nonparallel loading platens was necessary to reduce or eliminate the loss of prepared test specimens and produce a rugged protocol. The procedures for AMPT cyclic fatigue testing currently include two options for attaching nonparallel loading platens to the AMPT: shim and ball bearing. However, little to no background information exists on the relative merits or effects of these two methods. The shims or ball bearings are placed atop the topmost loading platen before tightening the topmost loading platen to the AMPT platen, as shown in figure 24 and figure 25. Considering these factors directly in the Plackett–Burman determination matrix was not possible since the factor (shims versus ball bearing) interacts with loading-platen parallelism. Therefore, the effect of these compensation mechanisms was evaluated by the research team in a separate experiment that informed proposed changes to the procedures for AMPT cyclic fatigue testing.
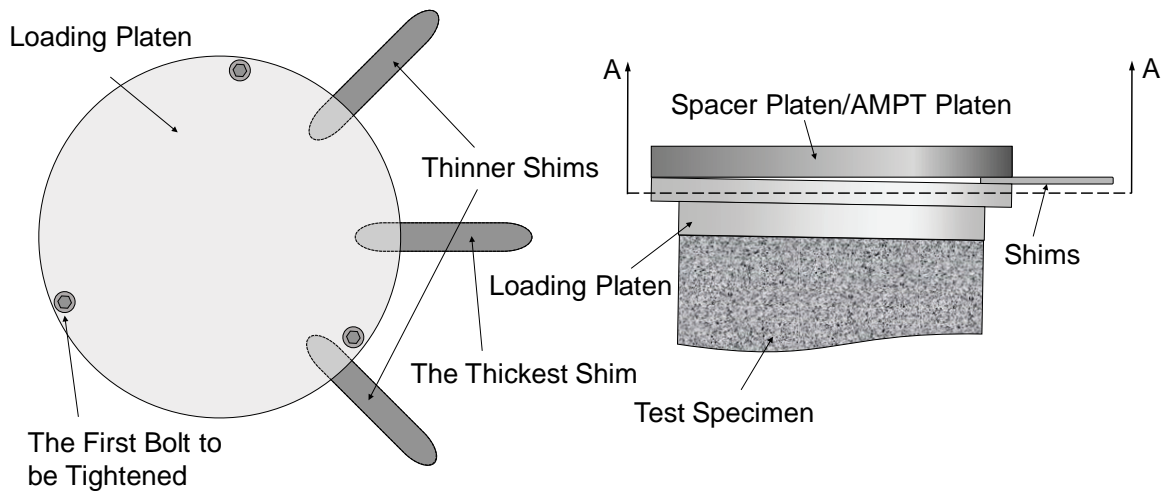
Source: FHWA.

Figure 24. Illustration. Schematic of shim placement.
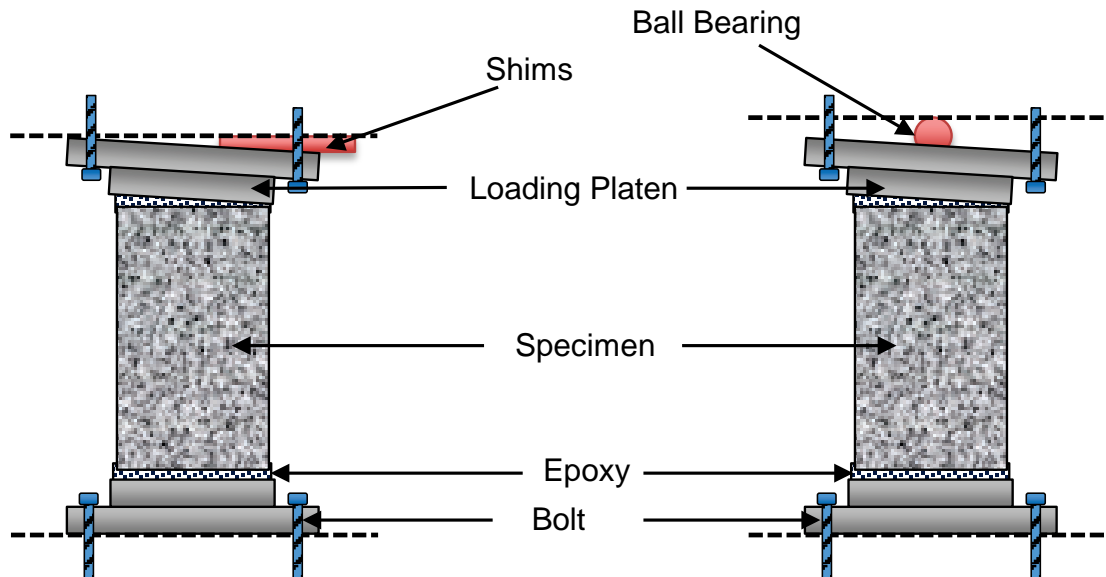
Figure 25. Illustration. Schematics of shim and ball bearing placements.

Attaching loading platens to the AMPT using the ball bearing introduces experimental complexity such that the ball bearing may introduce variability in the results of cyclic fatigue testing even when the loading platens are parallel. The standard does not prescribe any specific tightening procedure for users to follow when using the ball bearing, which may contribute to variability. Correspondingly, in this supplementary experiment, the strain that developed during the process of tightening the attachment bolts to the AMPT was assessed using parallel loading platens without the ball bearing, parallel loading platens with the ball bearing, nonparallel loading platens with the ball bearing, and nonparallel loading platens with the shim. In each case, AMPT operators followed a specific tightening pattern and process to better refine guidance in the standard. The glued specimens with nonparallel loading platens were prepared using an angle of 0.3 degrees.

Analytical Transfer Time Study

The research team conducted heat transfer analyses to determine the effects of laboratory and handling conditions on changes in specimen temperature that may take place during specimen transfer from an external chamber to the AMPT. Parameters specific to a given laboratory (e.g., laboratory temperature and humidity, test chamber temperature, transfer time, handling time) may affect temperature changes during specimen transfer. In this study, well-known and easily defined axisymmetric heat transfer functions, detailed in the Analytical Transfer-Time Study section in chapter 3, were coupled with experiments that monitored the temperature of specimens during transfer from an external conditioning chamber to the AMPT in the laboratory. The heat transfer analysis was performed using MATLAB®,[26] and the results were combined with the temperature error thresholds determined from the ruggedness evaluation to assess and improve guidance for temperature conditioning procedures.

Data-Quality Indicators Study

Refinement of DMR Limits Based on the Ruggedness Evaluation Results

DMR is used in AASHTO TP 107-18 and AASHTO TP 133-19 to characterize the potential difference between the specimens tested in the |*E**| test and the specimens used in the cyclic tension test. This value is also used to calculate pseudosecant modulus and damage parameters that are used in defining the damage-characteristic curve. The current specifications use the acceptable *DMR* range of 0.9 to 1.1. In this study, the research team established *DMR* limits based on the ruggedness-evaluation results. The ruggedness study revealed that three ruggedness experimental factors (i.e., the air-void content, temperature, and loading-platen parallelism) affect *DMR*. The team conducted a linear regression analysis using these three ruggedness factors to establish acceptable DMR limits.

Establishment of a Data-Quality Indicator to Ensure Proper Proportional, Integral, Derivative (PID) settings

AASHTO TP 107-18 and AASHTO TP 133-19 lack data-quality indicators to assess the efficacy and accuracy of the test. In their current form, these standards give users no objective means to determine whether systematic flaws in their experimental setup or equipment function exist. Natural variations and deviations are expected from these testing methods, as they are in any test. However, one error related to testing that has been found to be systematic and can affect the overall repeatability of individual cyclic fatigue tests is improper setup of the equipment's PID settings. The PID controller continuously calculates an error as the difference between a desired setpoint and a measured value and applies a correction based on proportional, integral, and derivative terms (denoted P, I, and D, respectively). Improper PID settings have two consequences: the actuator displacement signal may be nonsinusoidal, which will cause errors in the calculation of pseudostrain for the cyclic portion of the test, and the actuator displacement that is achieved in a test will not match the command signal, which will cause the experiment to run substantially longer (when the achieved signal is smaller than the command) or shorter (when the achieved signal is larger than the command) than expected. PID setting errors manifest from the very beginning of the testing, which means that the data collected in the initial five cycles can be used to determine whether errors exist. To develop an appropriate data-quality indicator in this study, data from the cyclic fatigue tests conducted as part of the ruggedness study (from both research laboratories) were combined with data acquired from the lead research laboratory and the participating partner agencies in the FHWA PRS shadow projects (i.e., DTFH61-13-C-00025 and DTFH61-14-D-00008).

The FHWA PRS shadow projects demonstrate to select agencies and contractors the process for accepting a pavement project and paying the contractor if PRS were used as a contract document. The PRS shadow projects also show the agency the ways in which the PRS may impact the agency's normal testing and volumetric-based acceptance operations.

To demonstrate the PRS, several activities (e.g., an AMPT workshop, onsite training, regular conference calls, and proficiency testing) were undertaken during the shadow project. The PRS shadow projects include data corresponding to six mixtures from four laboratories.

CHAPTER 3. RESULTS

OVERVIEW OF THE RUGGEDNESS-EVALUATION RESULTS

Dynamic Modulus Testing

Analyzing the results of the AMPT cyclic fatigue testing requires the use of $|E^*|$ test results. Air void is the only ruggedness-evaluation factor that affects $|E^*|$ testing. Therefore, the labs conducted $|E^*|$ tests at both air-void levels according to AASHTO T 378-17 and AASHTO TP 132-19 for large- and small-specimen geometries, respectively.[6,11] To best reflect the effect of a deviation from the target air-void content on the analysis results from the cyclic fatigue test, the $|E^*|$ test results at both air-void levels were analyzed together, and the averaged result was used in analyzing all results of the cyclic fatigue testing. The research team followed this approach because AASHTO TP 107-18 and AASHTO TP 133-19 specify that the average $|E^*|$ test results be coupled with the results of cyclic fatigue testing.[1,2]

The goal of the final allowable air-void tolerance in the standard for cyclic fatigue testing is to establish the maximum allowable deviation in air-void content from the target that results in statistically repeatable and reproducible results from cyclic fatigue testing. This target air-void content would be the same for both cyclic fatigue and $|E^*|$, and so the $|E^*|$ at the mid-point levels of air-void contents is most representative of what would happen in practice. This detail was not considered when originally setting up the plan for the ruggedness experiment, and the research team had performed $|E^*|$ tests at the same air-void levels as the cyclic fatigue tests. The $|E^*|$ test results are presented in the appendix. As expected, the results demonstrated higher $|E^*|$ for a given mixture at the lower air-void level than the higher air-void level. However, the difference in modulus between high and low air-void levels was approximately 19 percent on average and at most 37 percent. These differences are only slightly greater than the reproducibility standards and suggest that the averaging approach taken to create the representative modulus mastercurves for cyclic fatigue analysis likely provides an excellent estimate of the modulus at the mean air-void content.

Damage-Characteristic Curve and Failure-Criterion Results

Given that the ruggedness evaluation employed a partial factorial experimental design, visually inferring the effects of individual factors is difficult. However, the aggregated results of the ruggedness evaluation for each laboratory, mixture, and specimen geometry combination are presented in the following sections to show the overall variation introduced into the test results by varying the experimental factors.

Figure 26 through figure 37 show the results of each determination. Note that each determination included three tests that were replicated for a total of six cyclic fatigue tests. The damage-characteristic curves presented were generated by fitting the equation shown in figure 5 to the three tests comprising a given determination. In the D^R results presented, the height of the bars represents the average D^R value for the two replicates of a given determination and the error bars indicate the range in the D^R values between the determination replicates.

Large-Specimen Geometry

Figure 26 and figure 27 show the damage-characteristic curve results obtained from large-specimen testing of the 9.5-mm NMAS mixture by labs A and B, respectively. Lab B results demonstrate higher variability among the different determinations and determination replicates than lab A results. Trends in the determinations are difficult to infer visually as the variation in the damage-characteristic curves within determination replicates is comparable to the variation across determinations in many cases.
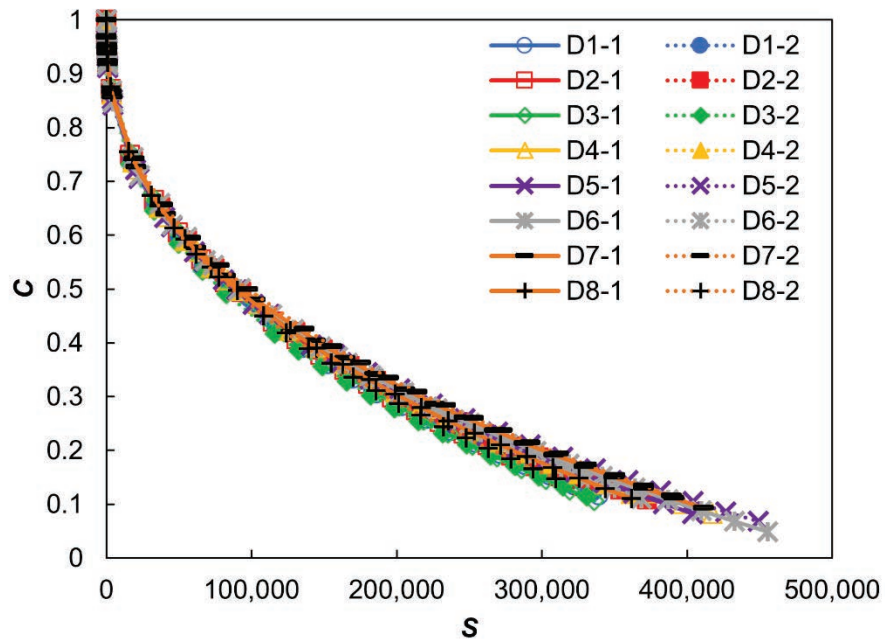
Source: FHWA.

Figure 26. Graph. Damage-characteristic curves for lab A large-specimen testing using 9.5-mm NMAS mixture.
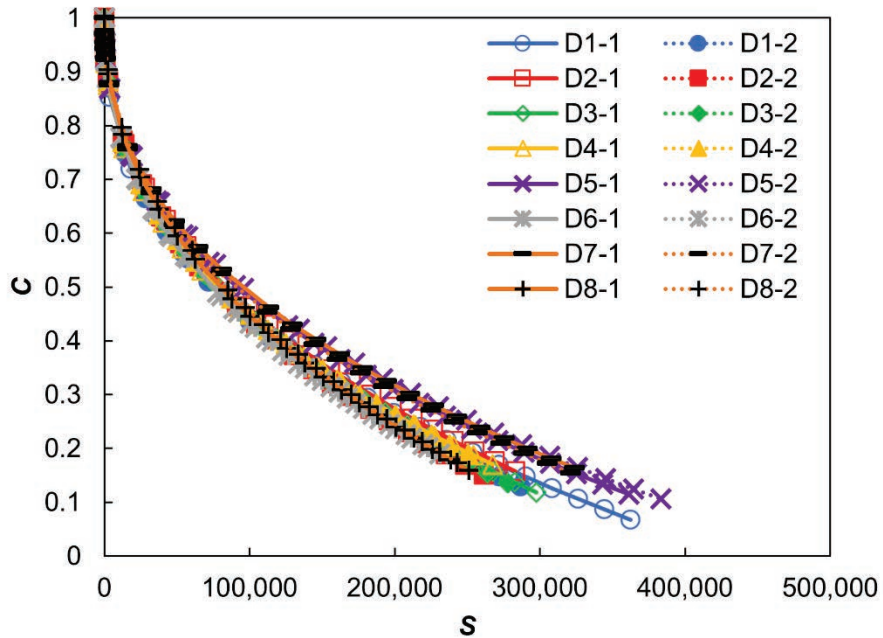
Figure 27. Graph. Damage-characteristic curves for lab B large-specimen testing using 9.5-mm NMAS mixture.

The D^R results of the large-specimen testing using the 9.5-mm NMAS mixture for labs A and B are shown in figure 28 and figure 29, respectively. The value of the bars represents the mean value of the two replicates of each determination, and the error bars represent the maximum and minimum values from the two determinations. Again, trends are difficult to infer visually given the relatively small variation in the D^R values across determinations.
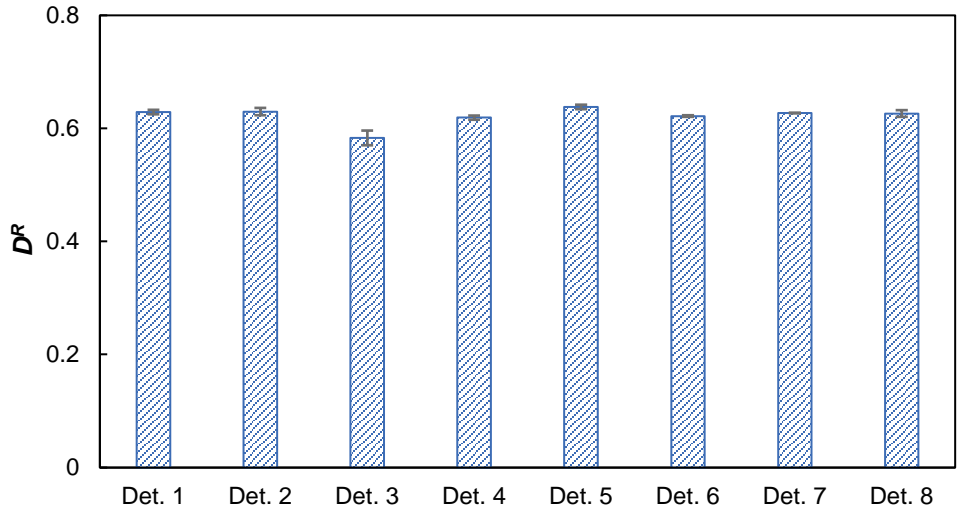
Figure 28. Graph. D^R failure criterion for lab A large-specimen testing using 9.5-mm NMAS mixture.
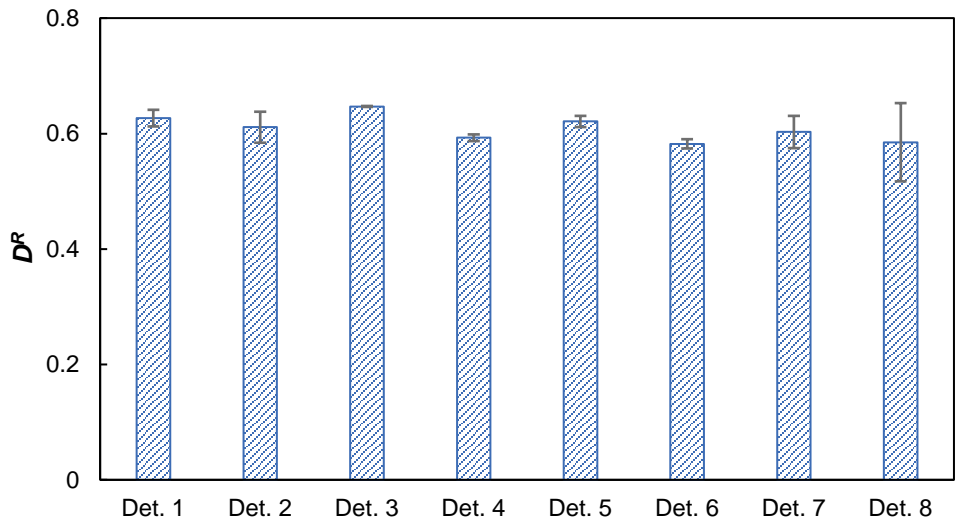
Figure 29. Graph. D^R failure criterion for lab B large-specimen testing using 9.5-mm NMAS mixture.

Figure 30 and figure 33 show the damage-characteristic curves and D^R failure-criterion results generated from ruggedness testing on the large specimens of the 25-mm NMAS mixture. Note that, unlike the 9.5-mm NMAS mixture, only lab A tested the 25-mm NMAS mixture. As shown in figure 30, the variability of damage-characteristic curves for the 25-mm mixture is much higher than that for the 9.5-mm mixture; this is not unexpected as higher NMAS mixtures

generally lead to higher variability in test results. Damage-characteristic curves from determinations 1 through 4 generally fall below the damage-characteristic curves from determinations 5 through 8. Furthermore, the damage-characteristic curves in determinations 1 through 4 generally exhibit less variability among determination replicates than determinations 5 through 8. Determinations 1 through 4 correspond to the high air-void level (i.e., 7 ± 0.5 percent) and determinations 5 through 8 correspond to the low air-void level (i.e., 5 ± 0.5 percent), suggesting air-void content may have a significant effect on the test results (both the values of the test results and the variability). The effects of other experimental factors are not visually discernable. Trends regarding determinations and experimental factors are not visually apparent in the D^R results presented in figure 33 given that the spans in the error bars are generally comparable to the variations across determinations.
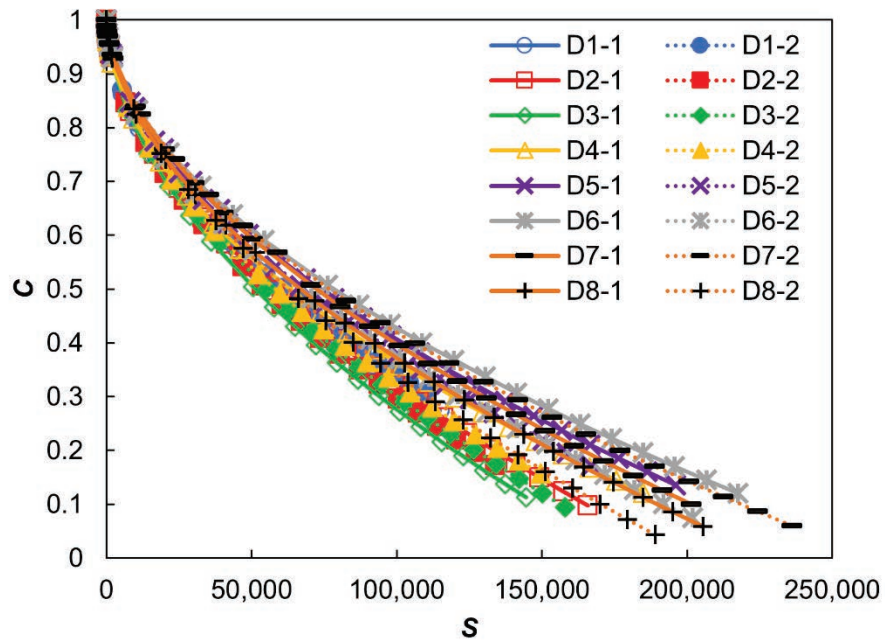
Source: FHWA.

Figure 30. Graph. Damage-characteristic curves for large-specimen testing using 25-mm NMAS mixture.
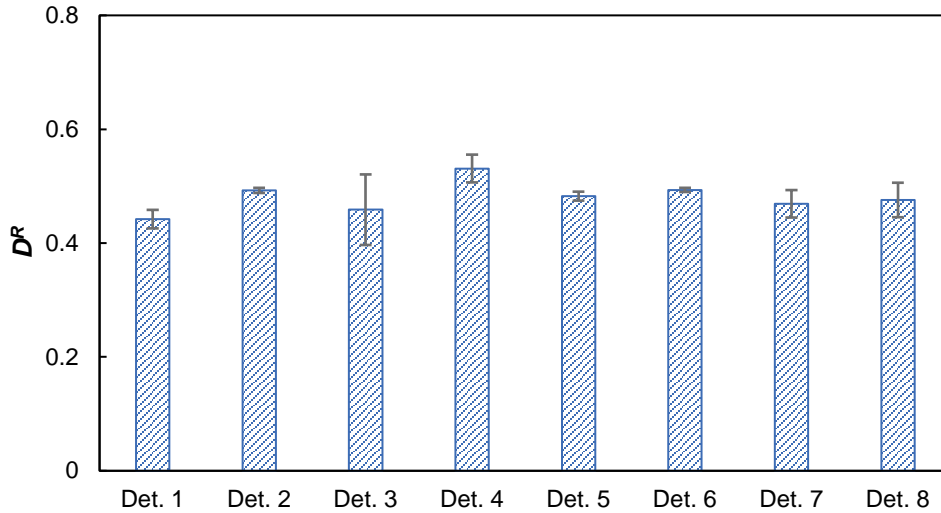
Figure 31. Graph. D^R failure criterion for large-specimen testing using 25-mm NMAS mixture.

Small-Specimen Geometry

Figure 32 and figure 33 show the damage-characteristic curves obtained from small-specimen testing using the 9.5-mm NMAS mixture by labs A and B, respectively. In lab A results, trends regarding determination are difficult to infer visually because the variation in the damage-characteristic curves within determination replicates is comparable to the variation across determinations. However, lab B damage-characteristic curves from determinations 1 through 4 tend to fall below the results of determinations 5 through 8, similar to the observed results for the large-specimen testing using a 25-mm NMAS mixture; this observation implies air-void content may affect the test results. The effects of other experimental factors are not visually discernable.
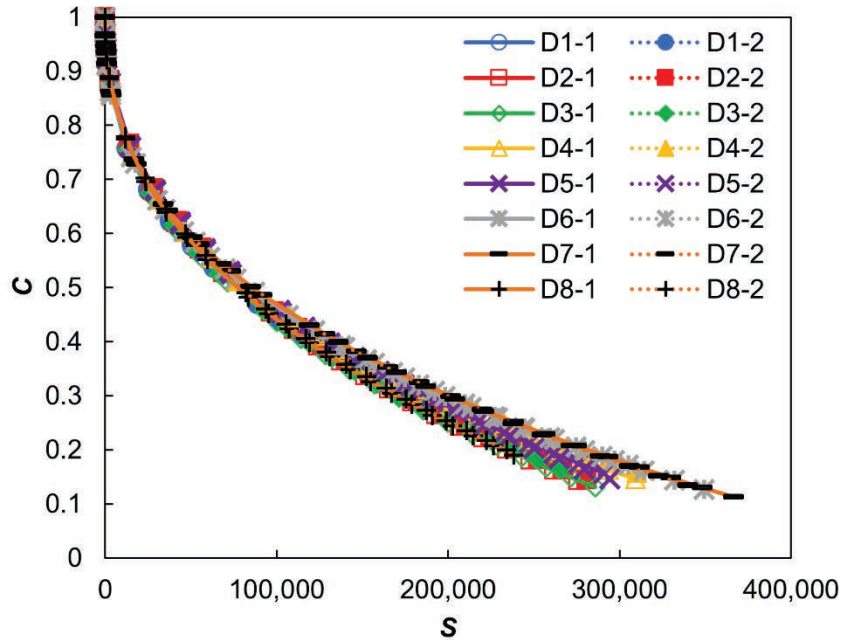
Source: FHWA.

Figure 32. Graph. Damage-characteristic curves for lab A small-specimen testing using 9.5-mm NMAS mixture.
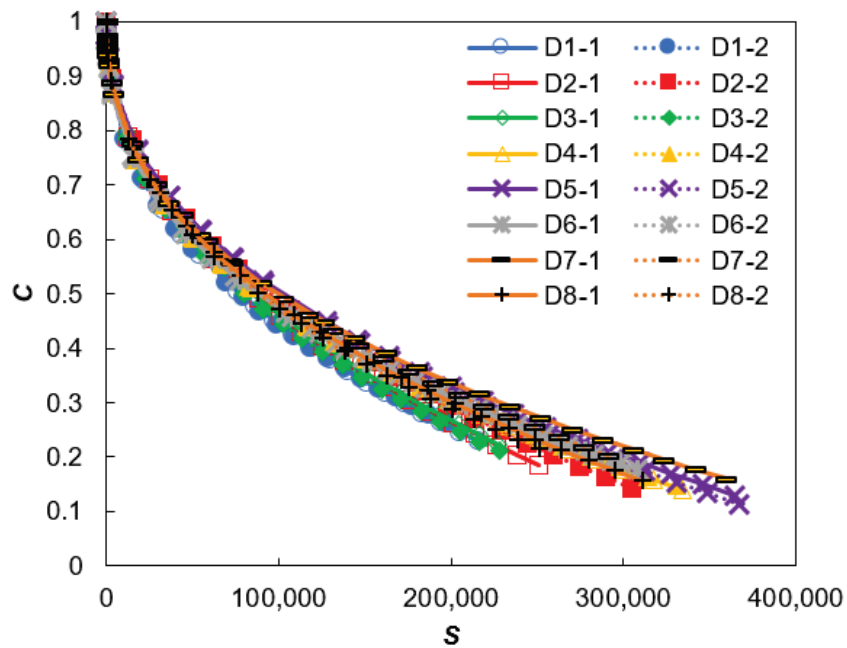
Source: FHWA.

Figure 33. Graph. Damage-characteristic curves for lab B small-specimen testing using 9.5-mm NMAS mixture.

35

The D^R results for the small-specimen 9.5-mm NMAS mixture from labs A and B are shown in figure 34 and figure 35, respectively. Lab A results show little variation among the different determinations, although lab B results show somewhat higher variability.
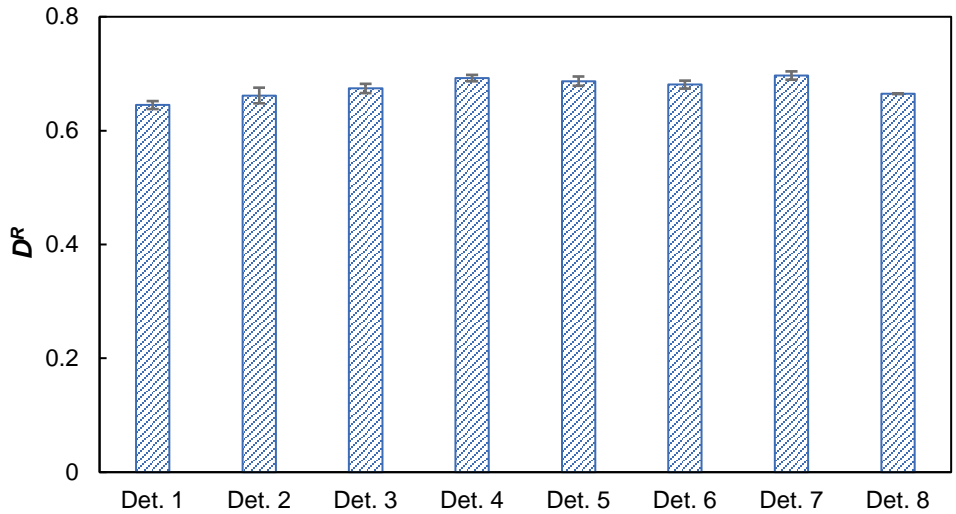
Figure 34. Graph. D^R failure criterion for lab A small-specimen testing using 9.5-mm NMAS mixture.
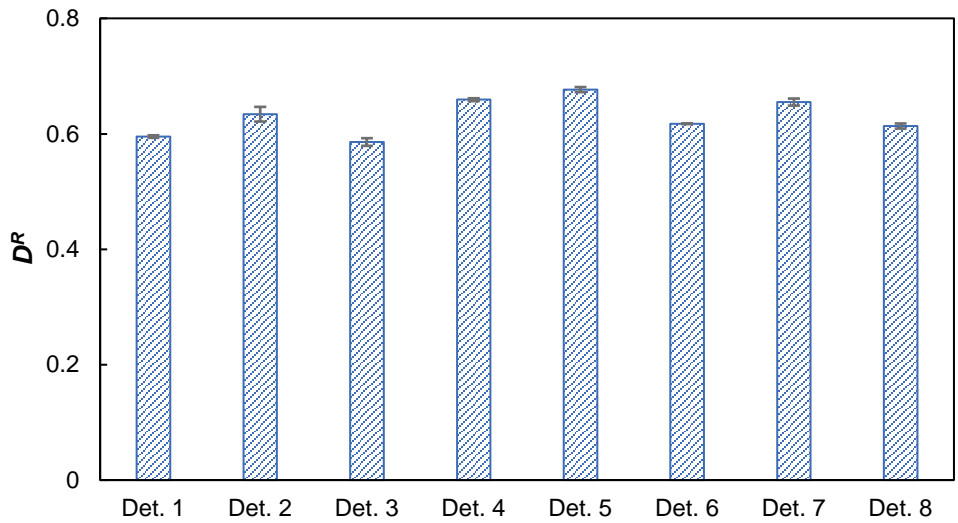
Figure 35. Graph. D^R failure criterion for lab B small-specimen testing using 9.5-mm NMAS mixture.

36

Figure 36 and figure 37 show the results for the damage-characteristic curve and D^R failure criterion in small-specimen testing using the 19-mm NMAS mixture, respectively. Lab A was the only lab to conduct experiments using the 19-mm NMAS mixture. The variation in both damage-characteristic curves and D^R failure criterion results for the 19-mm NMAS mixture is much higher than those for 9.5-mm NMAS mixture.
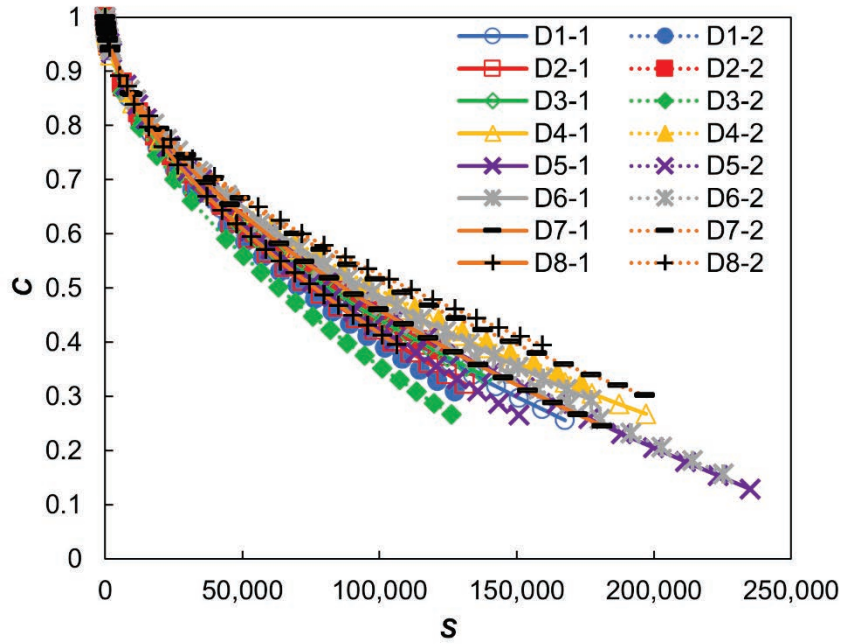
Source: FHWA.

Figure 36. Graph. Damage-characteristic curves for small-specimen testing using 19-mm NMAS mixture.
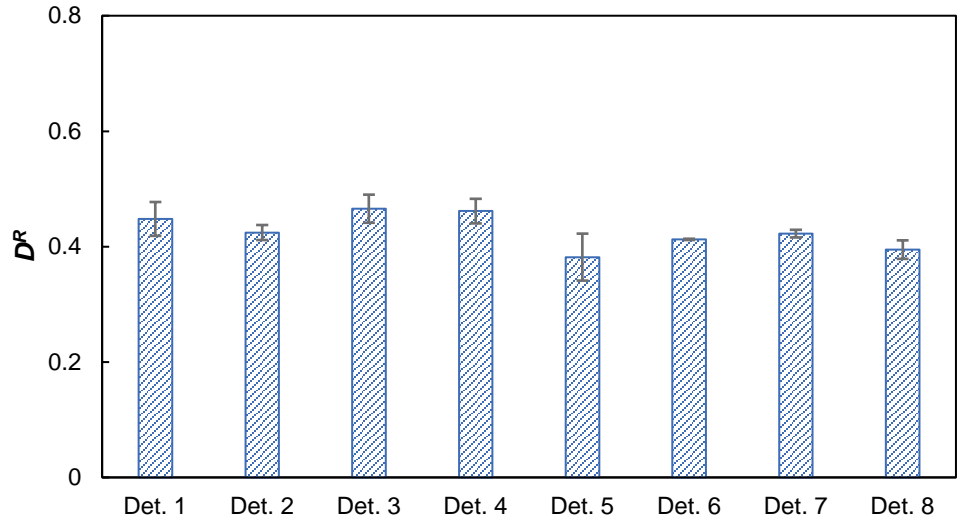
Figure 37. Graph. D^R failure criterion for small-specimen testing using 19-mm NMAS mixture.

End-Failure Occurrence

The preceding results correspond to tests that resulted in middle failure; that is, failure between the instrumented gauge points of the LVDT. However, some of the ruggedness experiments failed outside the gauge points (i.e., exhibited end failure). In these instances, tests were repeated using additional specimens until middle failures were obtained. Understanding the effects of the experimental factors on the occurrence of end failure can help guide refinement of the standards for AMPT cyclic fatigue testing to improve their ruggedness.

Table 8 through table 11 show the failure locations and number of end failures encountered in the ruggedness evaluations of the 9.5-mm mixture. In these ruggedness experiments, no measures to mitigate end failure (i.e., shims or ball bearing) were used. Each determination included three different strain levels, with the first strain level being the highest and the third being the lowest. As shown in table 8 and table 9, relatively few end failures were encountered in the ruggedness testing for the small specimen of the 9.5-mm mixture. Lab A experienced no end failures, and lab B experienced end failures in 5 out of the 48 tests. Note that, although it is not shown in this table, lab B reran the 9.5-mm mixture tests that initially resulted in end failures, and all the rerun tests yielded middle failures.

As shown in table 11, lab B encountered end failures in 9 out of the 48 large specimen ruggedness tests conducted using the 9.5-mm mixture. Most end failures occurred when the angled platens were used (i.e., determinations 2, 4, 6, and 8). In lab B's ruggedness evaluation using a large specimen, a loading-platen angle of 0.6 degrees was used as the nonparallel loading-platen condition. Given the relatively high occurrence of end failure, the loading-platen angle was reduced to 0.3 degrees in the ruggedness evaluations conducted by lab A using a large

specimen. Consequently, lab A encountered end failures in 3 out of 48 ruggedness tests using the large specimen of the 9.5-mm mixture, as shown in table 10.

Table 12 shows the failure locations (middle or end) for ruggedness-evaluation experiments conducted by lab A on the 25-mm large-specimen geometry. Note that a loading-platen angle for the angled condition of 0.3 degrees was used, and no end-failure mitigation strategies were employed. The results show that the ruggedness evaluation of the 25-mm mixture yielded more end failures than the 9.5-mm mixture (i.e., 18 end failures out of the original 48 tests).

Table 8. Failure locations when testing ruggedness in the 9.5-mm small specimen (lab A).

Replicate	Det. 1	Det. 2	Det. 3	Det. 4	Det. 5	Det. 6	Det. 7	Det. 8
1st rep. 1st strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
1st rep. 2nd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
1st rep. 3rd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 1st strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 2nd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 3rd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
No. of end failures	0	0	0	0	0	0	0	0

Det. = determination; rep. = replication.

Table 9. Failure locations when testing ruggedness in the 9.5-mm small specimen (lab B).

Replicate	Det. 1	Det. 2	Det. 3	Det. 4	Det. 5	Det. 6	Det. 7	Det. 8
1st rep. 1st strain	End	Middle	Middle	End	Middle	Middle	Middle	Middle
1st rep. 2nd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
1st rep. 3rd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	End
2nd rep. 1st strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 2nd strain	Middle	Middle	End	Middle	Middle	Middle	Middle	Middle
2nd rep. 3rd strain	End	Middle	Middle	Middle	Middle	Middle	Middle	Middle
No. of end failures	2	0	1	1	0	0	0	1

Det. = determination; rep. = replication.

Table 10. Failure locations when testing ruggedness in the 9.5-mm large specimen (lab A).

Replicate	Det. 1	Det. 2	Det. 3	Det. 4	Det. 5	Det. 6	Det. 7	Det. 8
1st rep. 1st strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
1st rep. 2nd strain	Middle	End	Middle	Middle	Middle	Middle	Middle	Middle
1st rep. 3rd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 1st strain	Middle	End	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 2nd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 3rd strain	End	Middle	Middle	Middle	Middle	Middle	Middle	Middle
No. of end failures	1	2	0	0	0	0	0	0

Det. = determination; rep. = replication.

Table 11. Failure locations when testing ruggedness in the 9.5-mm large specimen (lab B).

Replicate	Det. 1	Det. 2	Det. 3	Det. 4	Det. 5	Det. 6	Det. 7	Det. 8
1st rep. 1st strain	Middle	End	Middle	End	Middle	End	Middle	End
1st rep. 2nd strain	Middle	Middle	Middle	End	Middle	End	Middle	Middle
1st rep. 3rd strain	Middle	Middle	Middle	End	Middle	Middle	Middle	Middle
2nd rep. 1st strain	Middle	End	Middle	Middle	Middle	End	Middle	Middle
2nd rep. 2nd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 3rd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
No. of end failures	0	2	0	3	0	3	0	1

Det. = determination; rep. = replication.

Table 12. Failure locations when testing ruggedness in the 25-mm large specimen.

Replicate	Det. 1	Det. 2	Det. 3	Det. 4	Det. 5	Det. 6	Det. 7	Det. 8
1st rep. 1st strain	End	End	End	End	End	End	Middle	End
1st rep. 2nd strain	Middle	Middle	Middle	Middle	End	Middle	Middle	End
1st rep. 3rd strain	Middle	End	Middle	Middle	End	Middle	End	Middle
2nd rep. 1st strain	Middle	End	Middle	End	Middle	End	Middle	Middle
2nd rep. 2nd strain	Middle	Middle	Middle	End	Middle	Middle	Middle	Middle
2nd rep. 3rd strain	End	Middle	End	Middle	Middle	Middle	Middle	Middle
No. of end failures	2	3	2	3	3	2	1	2

Det. = determination; rep. = replication.

Unlike the 9.5-mm and 25-mm mixture tests, laboratory staff attempted to mitigate the amount of tensile strain that developed when attaching small specimens of the 19-mm mixture with nonparallel ends to the AMPT. This decision was based upon the observed occurrence of end failures when testing the 25-mm NMAS mixture with the large-specimen geometry combined with a high occurrence of end failure during pilot testing of the 19-mm mixture to establish the input strains for the ruggedness experiments. For the 19-mm nonparallel plate tests, shims were

placed between the top loading platen and AMPT platen. A supplementary investigation, presented in the End-Failure-Mitigation Study section in chapter 3, formed the basis of this decision and represented the practice that the labs followed during the ruggedness testing. Table 13 shows the end-failure results when shims were used for small-specimen testing using the 19-mm mixture. As shown in table 13, few end failures were encountered, which is attributed to the use of the shims to limit strain development upon attaching the specimen to the AMPT. End failures mostly occurred when the air-void content was low (i.e., determinations 5, 6, 7, and 8).

Table 13. Failure locations when testing ruggedness in the 19-mm small specimen.

Replicate	Det. 1	Det. 2	Det. 3	Det. 4	Det. 5	Det. 6	Det. 7	Det. 8
1st rep. 1st strain	Middle	Middle	Middle	Middle	Middle	Middle	End	End
1st rep. 2nd strain	Middle	Middle	Middle	Middle	Middle	End	Middle	Middle
1st rep. 3rd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 1st strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	End
2nd rep. 2nd strain	Middle	Middle	Middle	Middle	Middle	Middle	Middle	Middle
2nd rep. 3rd strain	Middle	Middle	Middle	End	Middle	Middle	Middle	Middle
No. of end failures	0	0	0	1	0	1	1	2

Det. = determination; rep. = replication.

STATISTICAL ANALYSIS OF THE RUGGEDNESS-EVALUATION RESULTS

Statistical analyses included *t*-tests and half-normal plots in accordance with ASTM E1169-14 and *F*-tests in accordance with ASTM C1067-12 to evaluate the significance of the experimental factors on the results of the ruggedness evaluation.[3,4] To evaluate the ruggedness test results as directly as possible, two metrics were used to evaluate the damage-characteristic curve results: the value of *C* at three specific *S* levels and the area above the damage-characteristic curve. The D^R failure criterion yields a single-point measurement; hence, the statistical analysis was applied directly to the test results. In addition, the fatigue-cracking index parameter, S_{app}, was used in the statistical analysis to evaluate the combined effects of the damage-characteristic curve and the failure criterion on a practical test outcome.

Table 14 shows the ruggedness factors identified as statistically significant using the different metrics and statistical tests at the material level. The factors in each cell of this table are ordered based on their *p* value, with the most statistically significant factors listed first. The critical *p* value and *F* value used to identify significant factors were 0.05 and 5.32, respectively. The research team selected these values to coincide with a 95-percent level of confidence. The detailed results of all statistical tests are presented in the appendix. Note that the different statistical tests led to the same outcome in terms of identifying significant experimental factors. Table 14 demonstrates that the statistically significant factors vary among the different mixtures, geometries, and laboratories. In addition, different test outcomes (e.g., damage-characteristic curve versus D^R failure criterion) are often affected by different experimental factors. However, as shown in figure 38, it is apparent that the air-void content, strain level, and test temperature are statistically significant more often than the other factors.

Table 14. Summary of significant factors.

Factor	Lab A 9.5-mm Small	Lab B 9.5-mm Small	Lab A 19-mm Small	Lab B 9.5-mm Large	Lab A 9.5-mm Large	Lab A 25-mm Large
C at low S value	Strain level Temp. Parallelism Air void Diameter	Strain level Temp.	Air void	Air void Temp. Strain level Height Diameter	—	Air void Temp.
C at medium S value	Strain level Air void Temp. Parallelism Height Rest period	Strain level Air void Temp.	—	Strain level Parallelism Air void	Air void Strain level Temp.	Air void Temp. Strain level
C at high S value	Temp. Strain level Air void Diameter Rest period Height	Strain level Air void	—	Strain level Parallelism Air void	Air void Strain level Temp.	Air void Temp. Strain level
Area above C versus S curve	Temp. Diameter Air void Strain level Rest period	Air void Strain level Diameter Height	—	Parallelism Strain level	Strain level Air void Temp.	Air void Strain level Diameter
D^R	Strain level Diameter Air void Height	Strain level Air void Rest period	Air void	—	Height Strain level Air void Rest period Diameter	—
S_{app}	Strain level Temp. Air void Diameter	Strain level Air void Diameter Rest period	—	Parallelism	Strain level Air void Height Temp.	Air void

—No significant factors.
Temp. = temperature.
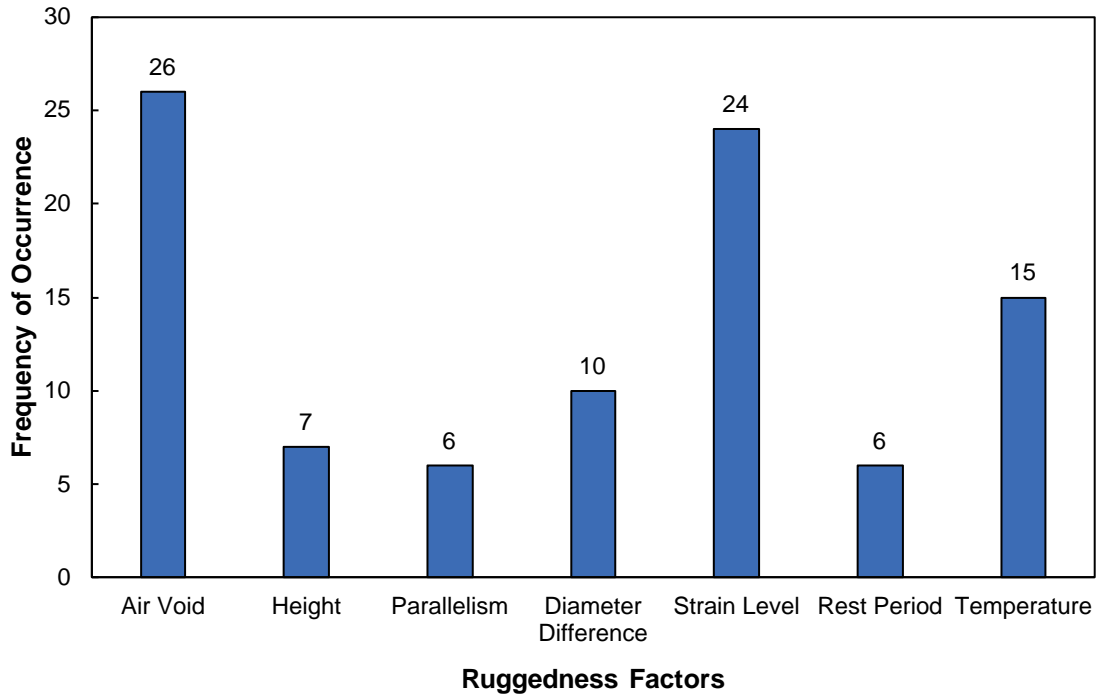
Source: FHWA.

Figure 38. Graph. Frequency of occurrence for each significant factor in statistical analyses.

The main effects of the statistically significant factors on the test outcomes are presented in figure 39 through figure 44. These graphs make it apparent that air voids and strain level generally have the most significant effect on test results. However, it is difficult to interpret the practical significance of the main effects using material-level parameters. Established tolerance thresholds for variation in the damage-characteristic curve are not currently available, making it difficult to draw practical inferences from the main effects presented in figure 39 through figure 42. Wang et al. evaluated the sensitivity of FlexPAVE simulations to variation in D^R values and found that when the differences in D^R values are less than 0.08, FlexPAVE predictions of fatigue damage do not differ significantly; this finding suggests all main effects of the experimental factors on D^R values shown in figure 43 are practically insignificant.[17] However, while the variation in D^R values alone may be practically insignificant, the main effects presented in figure 44 suggest that the effects of some experimental factors may be significant in terms of S_{app} values, which are calculated based on the combination of the damage-characteristic curve and D^R results. Acceptance thresholds for S_{app} values as a function of traffic level have been developed and proposed.[19] The S_{app} value thresholds among the different traffic levels differ by only six units in several cases. The main effects of the experimental factors in some instances approach or exceed a value of six units (e.g., air-void and strain-level factors assessed during the ruggedness evaluation conducted by lab B on the small specimen using the 9.5-mm mixture). These values suggest variations in the results of cyclic fatigue testing imposed by variations within the experimental ruggedness-factor levels could lead to different conclusions regarding the acceptable traffic level for which a mixture can be used based on S_{app} values. The proposed S_{app} thresholds were established based on the repeatability of the results of cyclic fatigue testing using an AMPT in a single laboratory, which may be smaller than the

43

repeatability and reproducibility statements ultimately allow. The ruggedness-evaluation levels represent extreme differences, thus making it difficult to assess the repeatability of S_{app} values using the ruggedness-evaluation results. For example, the difference in the target air-void content between the two experimental levels is 2.0 percent for the 9.5-mm mixture. The repeatability and reproducibility of S_{app} values with the imposed controls on the experimental factors defined from the ruggedness-evaluation results will be assessed in the interlaboratory study (ILS) and used to refine the S_{app} thresholds.
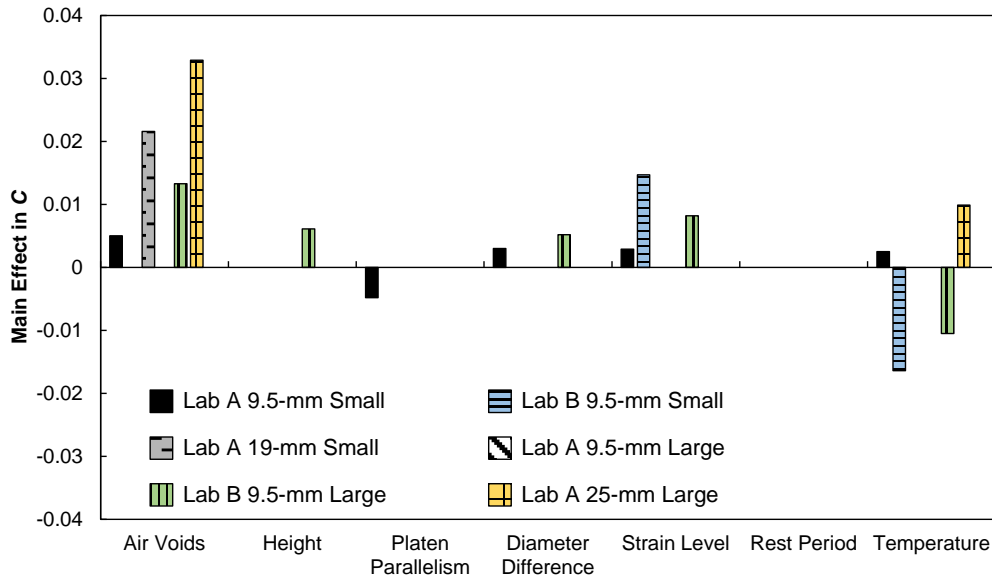
Source: FHWA.

Figure 39. Graph. Main effects for factors significantly affecting *C* at the low *S* value.
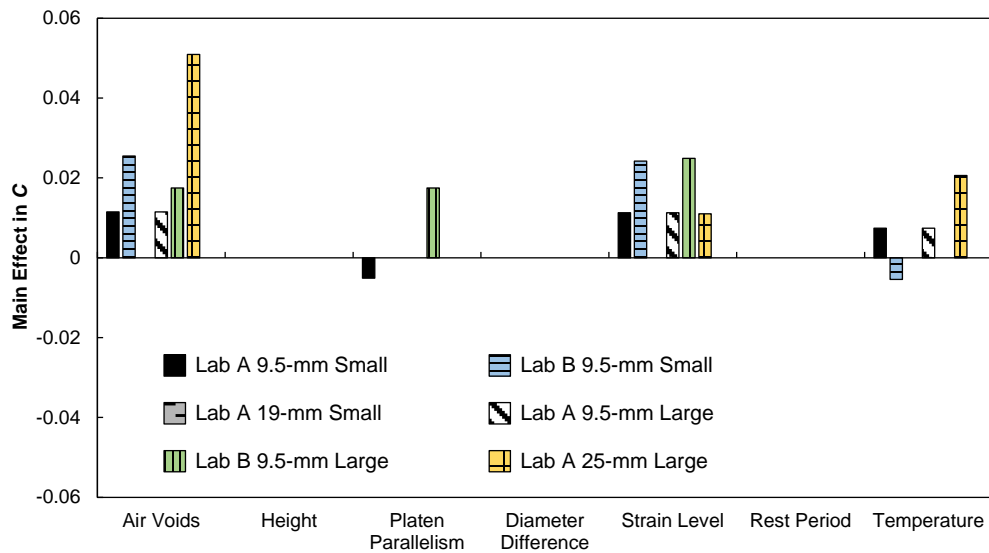
Source: FHWA.

Figure 40. Graph. Main effects for factors significantly affecting *C* at the medium *S* value.

44
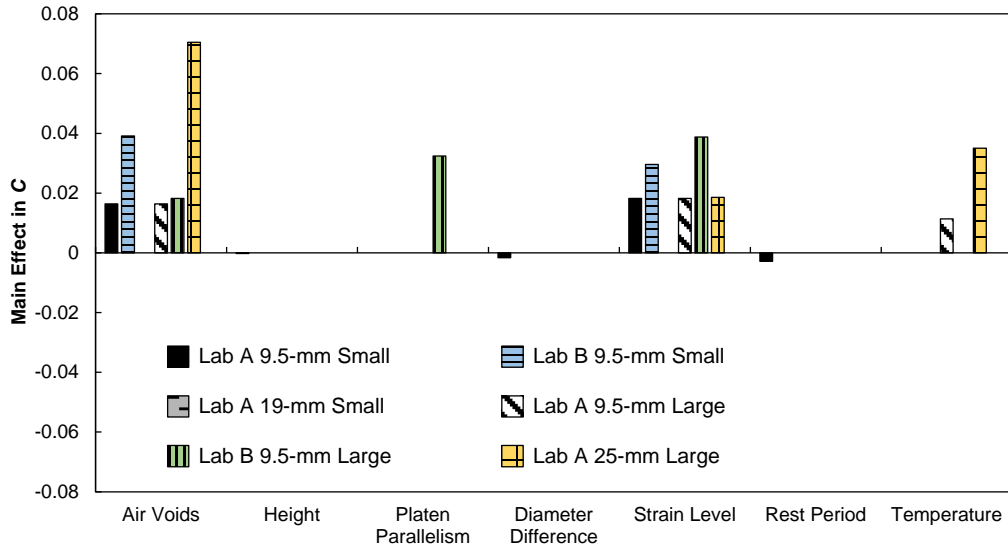
Source: FHWA.

Figure 41. Graph. Main effects for factors significantly affecting *C* at the high *S* value.
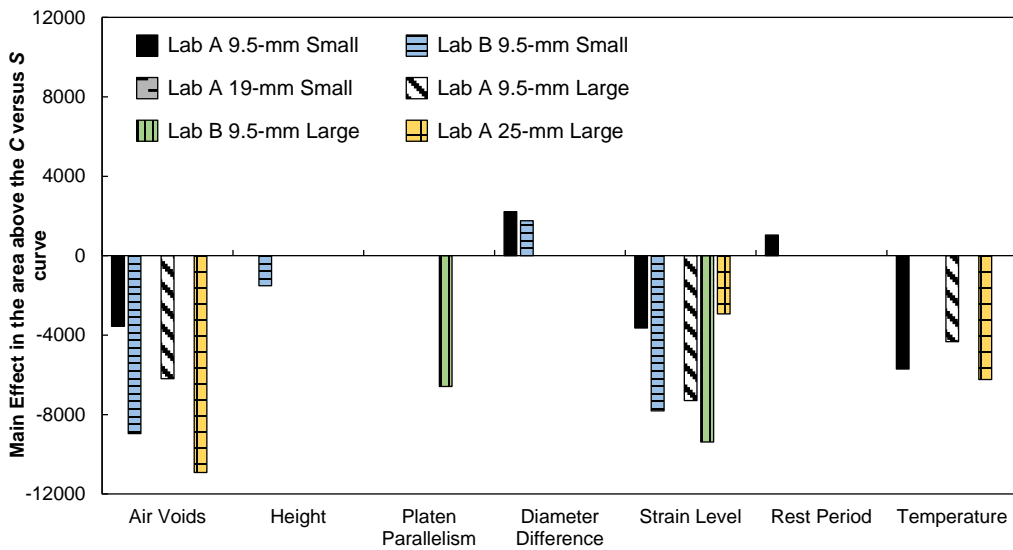
Source: FHWA.

**Figure 42. Graph. Main effects for factors significantly affecting the area above the
C versus *S* curve.**

45
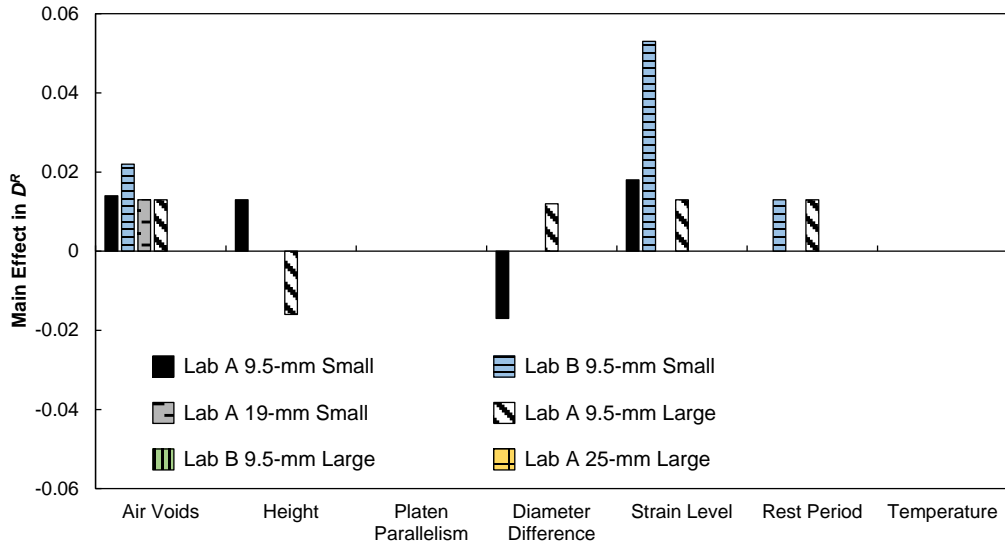
Source: FHWA.

Figure 43. Graph. Main effects for factors significantly affecting D^R.
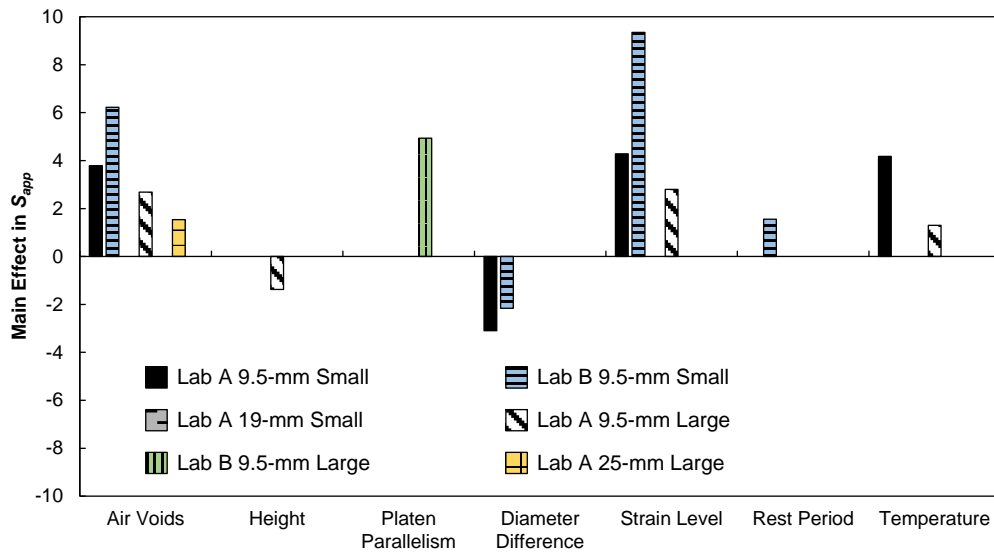
Source: FHWA.

Figure 44. Graph. Main effects for factors significantly affecting S_{app}.

THRESHOLD ANALYSIS OF THE RUGGEDNESS-EVALUATION RESULTS

Given the challenges in interpreting the practical significance of the experimental factors on material-level parameters, the research team used FlexPAVE simulation results to guide the establishment of tolerance thresholds for the experimental factors. FlexPAVE simulations made it possible to evaluate the combined effects of the differences in damage-characteristic curves and D^R failure criteria on predictions of *%Cracking* in the long-term performance of pavement—a practical and intuitive application of the test results. The team used a multivariate linear regression analysis of the FlexPAVE simulation results to establish the tolerance ranges for the

46

experimental factors. For each combination of mixture, laboratory, and test geometry, the ruggedness-evaluation results were fit to figure 23.

The research team first conducted a regression analysis on a determination-by-determination basis; that is, the results of the three tests comprising a given determination replicate were analyzed together as prescribed in AASHTO TP 107-18 and AASHTO TP 133-19 and used in the regression analysis.[1,2] Table 15 shows example results for the large-specimen testing in lab A using the 9.5-mm mixture. The comparison of *%Cracking* between measured and predicted values from these determinations is shown in figure 45. The results of the regression analysis conducted on a determination-by-determination basis yielded R^2 ranging from 0.69 to 0.98 for 9.5-mm mixtures, 0.74 for 25-mm mixture, and 0.51 for 19-mm mixture. The results are presented in the appendix.

Table 15. Sample results of regression analysis on determination-by-determination basis from lab A large-specimen testing using 9.5-mm NMAS mixture.

Determination	F1	F2	F3	F4	F5	F6	F7	FlexPAVE *% Cracking*	Regression Model *% Cracking*
D1	−1	−1	−1	1	1	1	−1	25.6	25.1
D2	−1	−1	1	1	−1	−1	1	22.6	20.5
D3	−1	1	−1	−1	1	−1	1	28.3	30.1
D4	−1	1	1	−1	−1	1	−1	20.4	20.9
D5	1	−1	−1	−1	−1	1	1	20.8	20.1
D6	1	−1	1	−1	1	−1	−1	20.3	19.5
D7	1	1	−1	1	−1	−1	−1	17.8	16.3
D8	1	1	1	1	1	1	1	19.8	21.5
D9	−1	−1	−1	1	1	1	−1	24.6	25.1
D10	−1	−1	1	1	−1	−1	1	18.5	20.5
D11	−1	1	−1	−1	1	−1	1	31.9	30.1
D12	−1	1	1	−1	−1	1	−1	21.4	20.9
D13	1	−1	−1	−1	−1	1	1	19.3	20.1
D14	1	−1	1	−1	1	−1	−1	18.8	19.5
D15	1	1	−1	1	−1	−1	−1	14.8	16.3
D16	1	1	1	1	1	1	1	23.2	21.5

D = determination; F = factor.
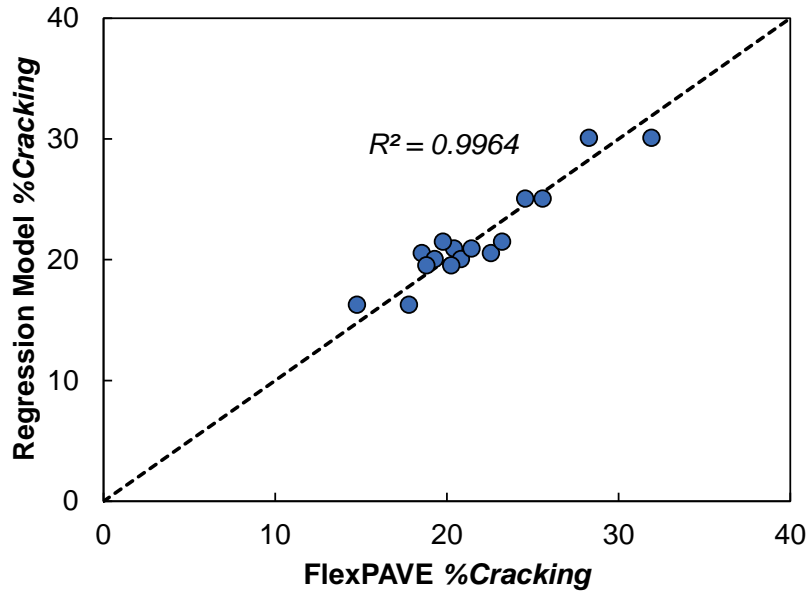
Source: FHWA.

Figure 45. Graph. Example results of predicted versus measured *%Cracking* on a determination-by-determination basis from lab A large-specimen testing using the 9.5-mm NMAS mixture.

The *p* values for the experimental factors resulting from the regression analysis conducted on a determination-by-determination basis are presented in table 16. Factors with *p* values below 0.05 were considered significant. The main effects of the factors that significantly affected the *%Cracking* predictions are presented in figure 46. Similar to the material-level analysis, the *p* value and main effect results of the FlexPAVE analyses indicate that air void and strain level generally have the largest effect on test results.

Table 16. *p* values for pavement-level statistical analyses.

Factor	Lab B 9.5-mm Small	Lab A 9.5-mm Small	Lab A 19-mm Small	Lab B 9.5-mm Large	Lab A 9.5-mm Large	Lab A 25-mm Large
Air void	<0.0000	0.0004	0.5669	0.2386	0.0009	0.0044
Specimen height	0.0503	0.2676	0.1531	0.5687	0.3676	0.6938
Loading-platen parallelism	0.0328	0.2781	0.2487	0.0098	0.0443	0.6745
Difference in diameter of loading platen and specimen	0.0264	0.0056	0.8964	0.6800	0.0915	0.3814
Strain level	<0.0000	0.0001	0.4793	0.0228	0.0012	0.1834
Rest period	0.7655	0.1082	0.7003	0.9837	0.7771	0.9657
Temperature	0.0429	0.0076	0.1067	0.0898	0.0248	0.0752
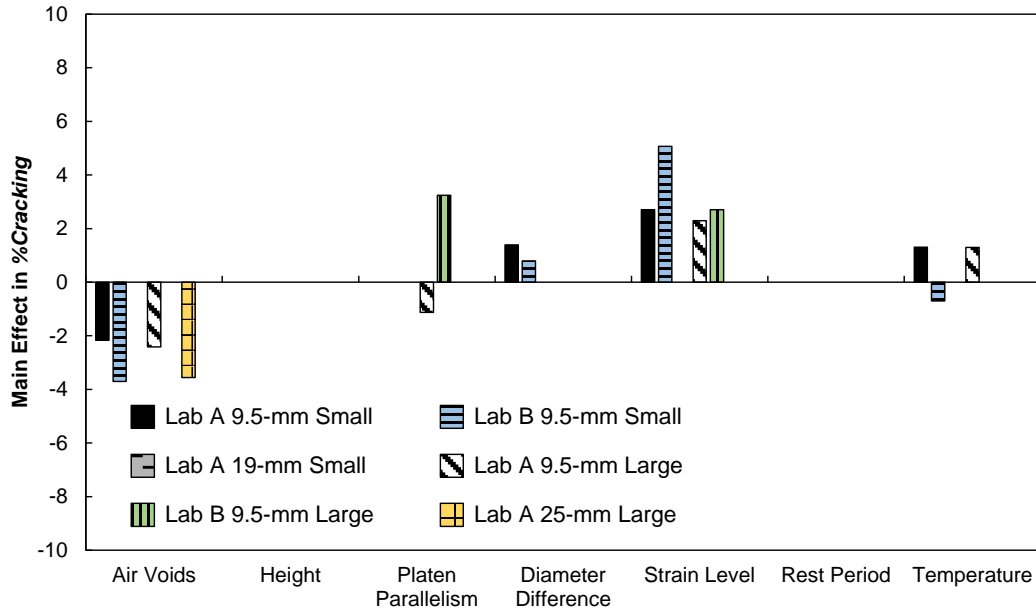
Source: FHWA.

Figure 46. Graph. Main effects for factors significantly affecting *%Cracking* results from FlexPAVE.

In the regression analysis, conducted on a determination-by-determination basis, the strain level factor was assigned a value of −1 or +1 according to the two levels used in the experimental plan. The outcome of the regression analysis to define tolerance ranges is thus an allowable range on this scale of −1 to +1. It is impossible to translate these results into an allowable N_f range without looking at the results of the individual tests for two reasons. First, N_f results from the ruggedness tests often deviated significantly from the target since N_f was controlled by the input strain levels, but specimen-to-specimen variability in N_f at a given strain level is high. Second, each determination consisted of three cyclic fatigue tests conducted at different input strain levels. Therefore, the research team also conducted a regression analysis based on individual specimen test results. In this analysis, the team used individual specimen test results to determine the damage-characteristic curve and D^R failure criterion, which were used as input for FlexPAVE simulations. The FlexPAVE simulation results were then used in the regression analysis for individual specimen test phases. In terms of four experimental factors (i.e., air void, specimen height, the difference in diameter between specimen and platen, and strain level), the results of the individual factors were scaled based on the −1 to +1 scale, as shown in table 17. Regarding strain level, in this analysis, the strain level factor values were assigned on the −1 to +1 scale based on the resultant N_f rather than the input strain level. The predictive capability of the regression model on individual specimen phases decreased substantially, as shown in figure 48. Therefore, this regression analysis was not used to define tolerance ranges for the significant experimental factors. Consequently, the ruggedness-evaluation results could not be used to inform refinement of the allowable range in N_f in AASHTO TP 107-18 and AASHTO TP 133-19.

49

Table 17. Example results of regression analysis on individual specimen basis from large-specimen testing by lab A using the 9.5-mm NMAS mixture.

Det.	F1	F2	F3	F4	F5	F6	F7	FlexPAVE *%Cracking*	Regression Model *%Cracking*
D1-1	−0.50	−0.82	1	0.96	0.28	1	−1	22.7	22.9
D1-2	−0.50	−0.92	−1	0.96	0.58	1	−1	23.0	24.0
D1-3	−0.80	−1.06	−1	0.96	1.22	1	−1	26.1	27.1
D1-4	−0.80	−0.96	−1	0.98	0.29	1	−1	27.6	23.6
D1-5	−0.50	−1.14	−1	0.96	0.49	1	−1	22.9	23.5
D1-6	−0.80	−0.92	−1	0.96	0.58	1	−1	24.5	24.2
D2-1	−1.00	−0.94	1	0.96	−0.60	−1	1	23.4	21.5
D2-2	−0.80	−1.06	1	0.96	−0.30	−1	1	26.1	22.1
D2-3	−0.90	−0.94	1	0.98	0.19	−1	1	23.1	24.2
D2-4	−0.60	−0.98	1	0.98	−0.20	−1	1	20.8	22.0
D2-5	−0.50	−1.06	1	0.96	0.11	−1	1	17.0	22.9
D2-6	−0.70	−0.99	1	0.98	0.19	−1	1	22.1	22.5
D3-1	−0.60	0.88	−1	−1.02	−0.10	−1	1	31.0	26.9
D3-2	−0.50	1.10	−1	−1.02	0.27	−1	1	23.1	28.1
D3-3	−0.60	0.96	−1	−1.04	0.95	−1	1	28.0	30.9
D3-4	−0.60	1.06	−1	−1.04	−0.18	−1	1	27.5	26.7
D3-5	−0.70	0.98	−1	−1.04	0.36	−1	1	24.4	28.9
D3-6	−0.70	0.98	−1	−1.04	0.53	−1	1	38.9	29.6
D4-1	−0.90	1.04	1	−1.04	−0.65	1	−1	16.2	21.1
D4-2	−0.70	1.16	1	−1.02	−0.68	1	−1	19.0	20.5
D4-3	−0.60	0.94	1	−1.04	−0.24	1	−1	23.7	21.8
D4-4	−0.50	0.82	1	−1.02	−0.90	1	−1	17.3	18.9
D4-5	−0.70	1.02	1	−1.04	−0.44	1	−1	27.1	21.3
D4-6	−0.80	0.94	1	−1.04	−0.09	1	−1	21.4	22.8
D5-1	1.00	−0.98	−1	−1.04	−0.88	1	1	14.7	19.0
D5-2	1.00	−1.14	−1	−1.04	−0.84	1	1	27.1	19.1
D5-3	0.60	−0.98	−1	−1.02	−0.24	1	1	23.7	22.5
D5-4	1.10	−1.14	−1	−1.04	−0.46	1	1	18.7	20.2
D5-5	0.70	−0.96	−1	−1.02	−0.39	1	1	17.3	21.6
D5-6	0.80	−1.08	−1	−1.04	−0.27	1	1	21.0	21.8
D6-1	0.90	−0.98	1	−1.04	−0.10	−1	−1	16.5	16.7
D6-2	1.00	−0.98	1	−1.04	0.32	−1	−1	12.9	18.0
D6-3	1.00	−1.08	1	−1.02	1.71	−1	−1	28.5	23.2
D6-4	1.00	−0.92	1	−1.02	−0.31	−1	−1	15.1	15.6
D6-5	0.90	−0.98	1	−1.04	0.20	−1	−1	18.4	17.8
D6-6	0.80	−1.14	1	−1.04	0.90	−1	−1	22.1	20.6
D7-1	0.90	0.98	−1	0.96	−0.76	−1	−1	15.6	16.1
D7-2	1.30	0.82	−1	0.96	−0.62	−1	−1	15.9	15.5
D7-3	1.00	1.02	−1	0.98	−0.20	−1	−1	20.2	18.0
D7-4	1.00	1.06	−1	0.96	−0.96	−1	−1	15.2	15.2
D7-5	1.00	1.16	−1	0.98	−0.60	−1	−1	14.8	16.5
D7-6	1.30	0.98	−1	0.96	−0.37	−1	−1	15.1	16.5
D8-1	1.00	0.92	1	0.96	−0.29	1	1	21.1	19.2
D8-2	0.80	0.96	1	0.96	0.35	1	1	18.2	22.2
D8-3	0.90	1.02	1	0.98	1.04	1	1	21.1	24.6

Det.	F1	F2	F3	F4	F5	F6	F7	FlexPAVE %Cracking	Regression Model %Cracking
D8-4	0.80	1.12	1	0.96	−0.53	1	1	23.6	19.0
D8-5	1.00	0.84	1	0.96	0.24	1	1	24.1	21.2
D8-6	0.90	0.98	1	0.96	0.56	1	1	23.3	22.8

D = determination; F = factor.
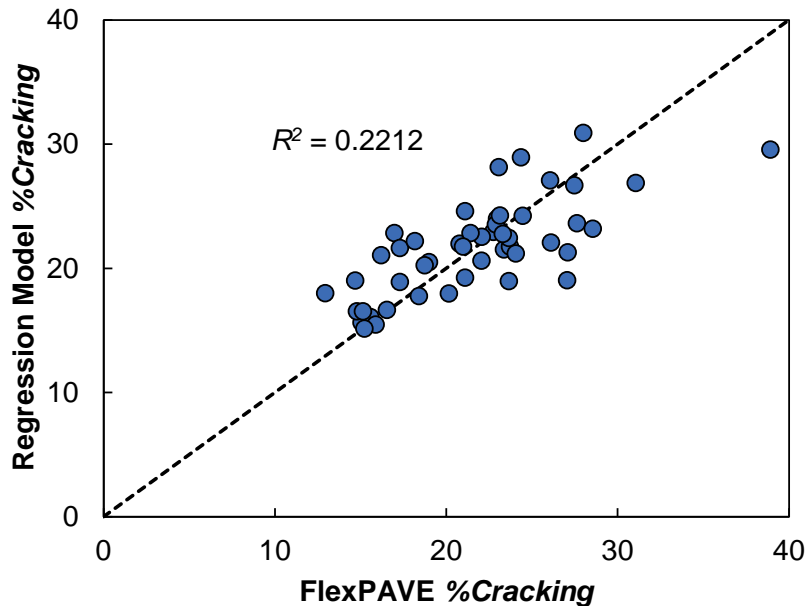
Source: FHWA.

Figure 47. Graph. Example results of predicted versus measured *%Cracking* on individual-specimen basis from lab A large-specimen testing using the 9.5-mm NMAS mixture.

To establish the tolerance thresholds for the other significant experimental factors, the regression analysis conducted on a determination-by-determination basis was rerun using only significant experimental factors. The resultant model coefficients were used to estimate the range in values for each factor that will keep the error in predicted cracking below a tolerance threshold of ±10 percent of the value of the predicted *%Cracking* when all experimental factors are at an ideal experimental condition. The ideal experimental condition is the one that hits the specified target value for all experimental factors (i.e., if height is specified as 110 ± 2.5 mm, then the ideal condition for height is 110 mm). As such, the ideal experimental condition is defined in the regression model by assigning a value of zero for each factor (X_{1-7} from figure 23) except for loading-platen parallelism, which has a value of −1 (perfectly parallel) in the ideal condition. The resultant tolerance thresholds for large- and small-specimen geometries are shown in table 18 and table 19, respectively. Except for air-void content and loading-platen angle, the calculated tolerance ranges are wider than the experimental levels in the ruggedness evaluation. The results, therefore, suggest that the current guidance in the testing standards is rugged because, except for air-void content, the experimental levels were selected based on the extremes currently permitted in AASHTO TP 107-18 and AASHTO TP 133-19. Because the regression analysis could not be used to define tolerance ranges for the strain-level factor, any variability associated with the allowable N_f range in the testing standards (i.e., 2,000 to 80,000 cycles) will be embedded into

the repeatability and reproducibility statements that will be established using interlaboratory studies. Since the calculated tolerance ranges for air-void content and loading-platen parallelism are narrower than the experimental levels in the ruggedness evaluation, the calculated ranges for those two factors need to be considered to establish the tolerance ranges for each factor. To address a lack of loading-platen parallelism, strategies to mitigate the development of strain upon attaching the specimen to the AMPT are presented in the next section.

Table 18. Calculated tolerance ranges in the ruggedness evaluation for large specimens.

Factor	Ruggedness Experimental Condition	Lab A 9.5-mm NMAS Mixture	Lab B 9.5-mm NMAS Mixture	Lab A 25-mm NMAS Mixture
Air void	±1.0%	±0.9%	*	±0.8%
Height	±5 mm	*	*	*
Loading-platen parallelism	0° to 0.3°/0.6°	0° to 0.3°	0° to 0.1°	*
Difference in diameter of loading platen and specimen	−4 to +7 mm	*	*	*
Rest period	5 to 20 min	*	*	*
Temperature	±2°C	±3.5°C	*	*

*Factor not statistically significant.

Table 19. Calculated tolerance ranges in the ruggedness evaluation for small specimens.

Factor	Ruggedness Experimental Condition	Lab A 9.5-mm NMAS Mixture	Lab B 9.5-mm NMAS Mixture	Lab A 19-mm NMAS Mixture
Air void	±1.0%	±0.7%	±0.4%	*
Height	±2.5 mm	*	*	*
Loading-platen parallelism	0° to 0.6°	*	0° to 0.6°	*
Difference in diameter of loading platen and specimen	−2 to +3 mm	−1.8 to +3.2 mm	−4 to +5 mm	*
Rest period	5 to 20 min	*	*	*
Temperature	±2°C	±2.3°C	±4.1°C	*

*Factor not statistically significant.

END-FAILURE-MITIGATION STUDY

The research team sought to clarify the need for a ball bearing or shims to reduce or eliminate the occurrences of end failures and improve the overall ruggedness of the protocols. The study used the 25-mm NMAS mixture because end failure was most prevalent with this mixture during the ruggedness study. The first step in this investigation was to measure the strain that develops in the specimen when it is attached to the AMPT. The case where the loading-platen angle was 0.3 degrees (i.e., consistent with the ruggedness experiments) was used for three scenarios: no mitigation (i.e., directly tightening the specimen's angled loading platen and to the AMPT), ball bearing, and shims. In all cases, the research team adhered to the following tightening procedure when attaching the specimen to the AMPT:

1. Set up the specimen in the AMPT chamber and tighten the bottom loading platen with three bolts.
2. Raise the actuator and finger tighten three bolts to the top loading platen.
3. Apply a 30-degree rotation to each bolt, starting with the bolt on the opposite side from the gap and moving around the specimen.
4. Repeat step 3 until all three bolts are tightened (approximately six to eight rounds).

For this process, a standard hex wrench (12 by 3 cm) aligned in the orientation that maximized the torque advantage was used. Strains were monitored during bolt tightening using LVDTs mounted on the specimen. Three LVDTs colored green, blue, and black were used, with the green LVDT being aligned with maximum gap. Figure 48 shows the onspecimen strain development when the specimen with the angled loading platen was attached to the AMPT without a ball bearing or shim. In this case, the strain development is very high. Figure 49 shows the onspecimen strain development when the specimen with the angled loading platen was attached to the AMPT with a ball bearing. In this case, the ball bearing greatly diminishes the strain development during attachment to the AMPT, but note the results with the ball bearing are highly contingent on the tightening pattern. Figure 50 shows the onspecimen strain development while attaching the loading platen to the AMPT using a ball bearing when the loading platens are parallel. The development of strain upon attaching the specimen to the AMPT is comparable with parallel and nonparallel loading platens when the ball bearing is used.
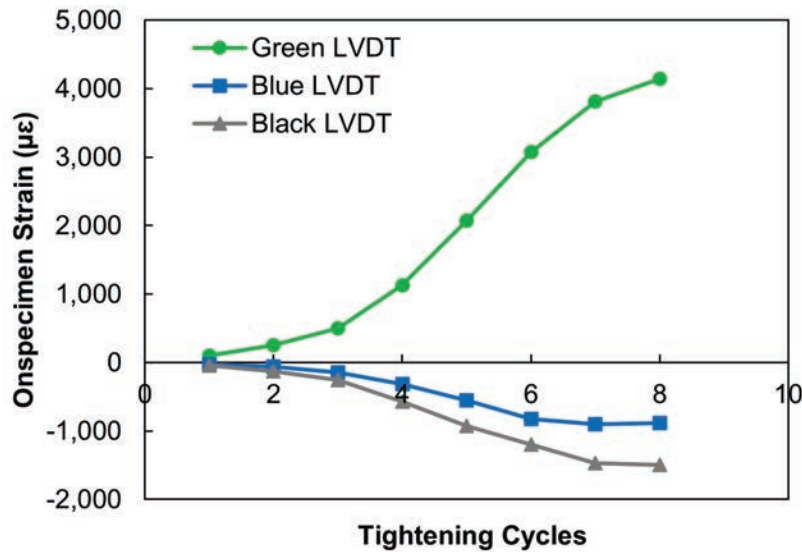
Source: FHWA.

Figure 48. Graph. Strain development during tightening with gap and loading-platen angle of 0.3 degrees.
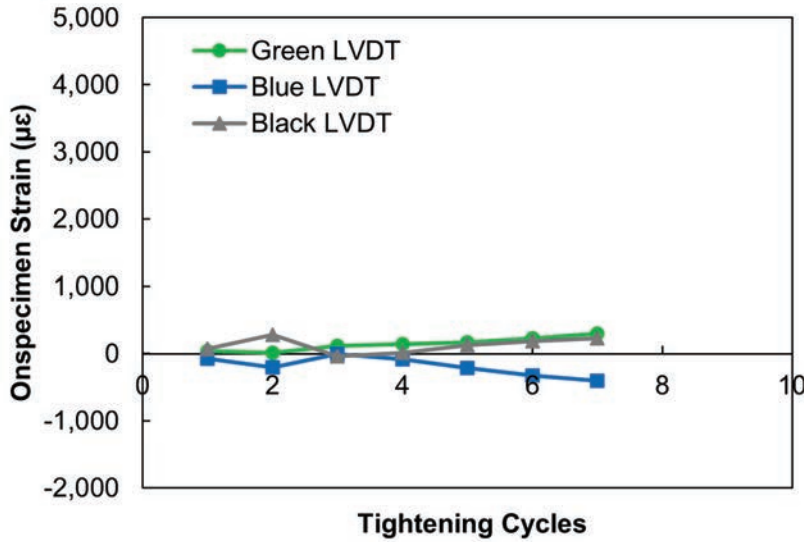
Source: FHWA.

Figure 49. Graph. Strain development during tightening with loading-platen angle of 0.3 degrees and ball bearing.
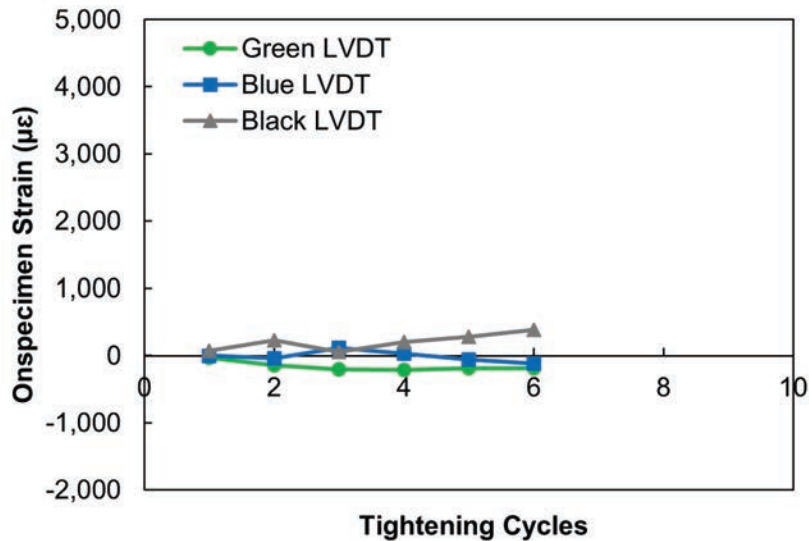
Source: FHWA.

Figure 50. Graph. Strain development during tightening with parallel loading platens and ball bearing.

Figure 51 shows onspecimen strain development when the specimen with the angled loading platen was attached to the AMPT and shims were used to fill the gaps between the upper loading platen and AMPT. The shims mitigate the development of strain during the process of attaching the specimen to the AMPT. Note that the vertical axis scale used in figure 51 is much smaller than that in figure 48 through figure 50. Figure 52 contains the same results for onspecimen strain development when the ball bearing is used with the angled plate as those shown in figure 49, but the graph uses the same scale as figure 51. It shows that shim use is more effective at mitigating strain development than the ball bearing.

54
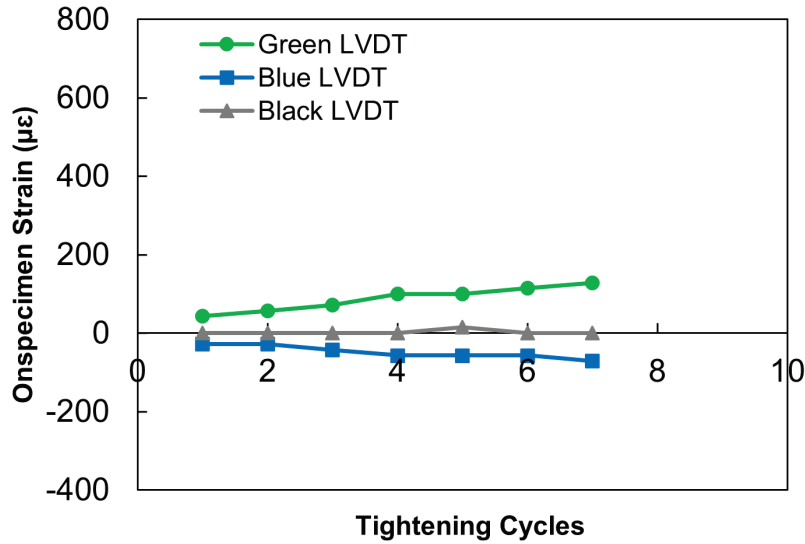
Source: FHWA.

Figure 51. Graph. Strain development during tightening with gap and with shim.
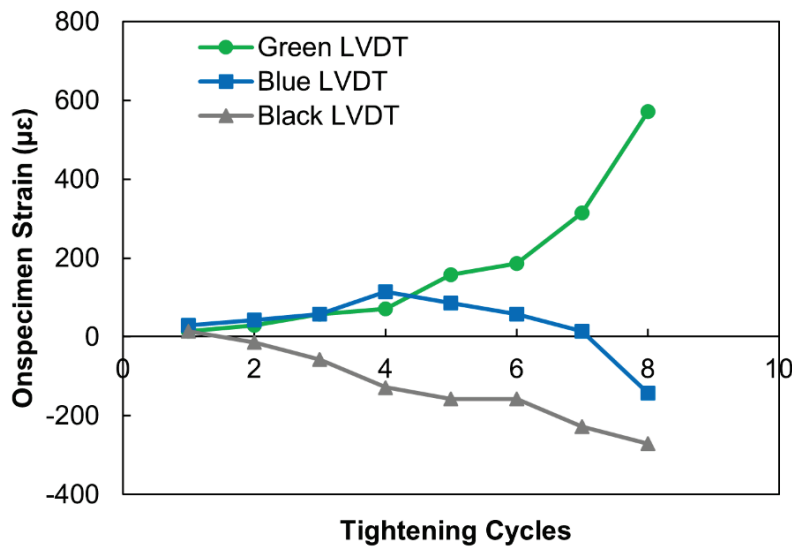
Source: FHWA.

Figure 52. Graph. Strain development during tightening with gap and ball bearing.

Lab A conducted cyclic fatigue testing on a large specimen using the 25-mm NMAS ruggedness mixture to further evaluate the abilities of the shims and ball bearing to mitigate end-failure occurrence. In these experiments, fatigue-test conditions were selected based on a combination of strain level and air void that resulted in repeated end failures in ruggedness testing. These conditions were combined with other ideal test conditions, as follows:

- Air void: 6 percent.
- Specimen height: 130 mm.
- Loading-platen angle: 0.3 degrees.

- Diameter: 100 mm.
- Strain Level: 220 microstrains (this is also the highest strain level in the ruggedness testing for 25-mm mixture).
- Rest period: 5 min.
- Temperature: 21°C.

Note that the strain level chosen was the highest strain level used in the ruggedness testing, and this testing resulted in 100 percent end failure. Two tests were conducted under these conditions: one with the ball bearing and one with shims. Both tests resulted in middle failure, indicating that both the use of shims and the ball bearing aid in mitigating end failure. However, given that the shims better limit strain development during attachment to the AMPT, using shims rather than ball bearings is recommended. Additional experiments to further evaluate the use of shims to mitigate end failure were conducted with loading-platen angles of 0.6 and 1.2 degrees with the shimming procedure. In the additional experiments, both 0.6- and 1.2-degree tests resulted in middle failures. The damage-characteristic curve and failure-criterion results are shown in figure 53 and figure 54. The results demonstrate very good agreement between the results of specimens with different loading-platen angles, demonstrating that the shims do not increase specimen-to-specimen variability.
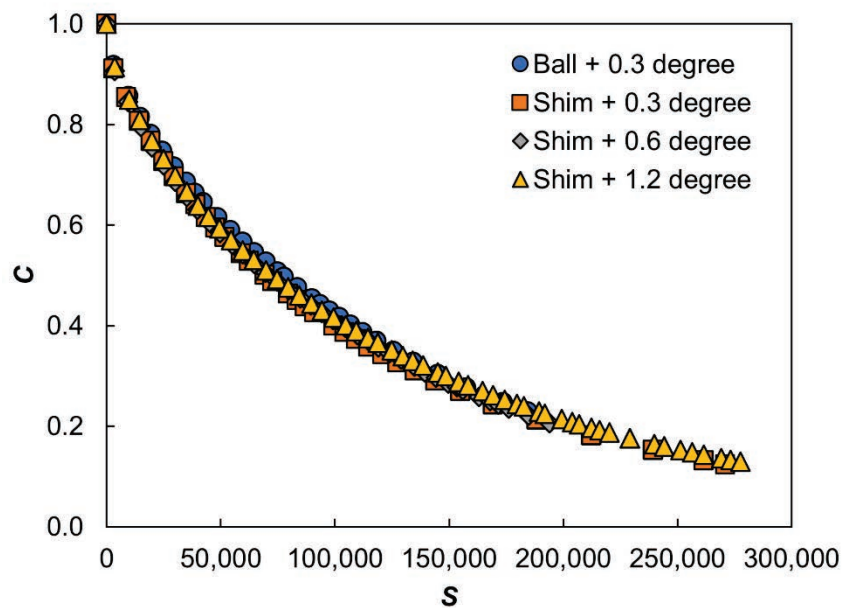
Source: FHWA.

Figure 53. Graph. Damage-characteristic curve results.
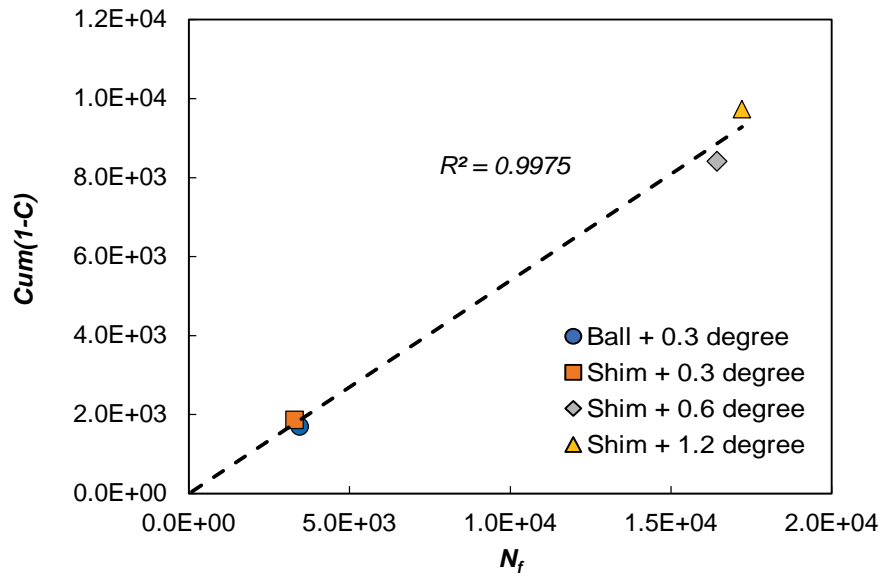
Source: FHWA.

Figure 54. Graph. Failure-criterion results.

The ruggedness-evaluation results for 19-mm NMAS mixture showed that use of shims up to the loading-platen angle of 0.6 degrees has no significant effect on the results of cyclic fatigue testing using a small specimen. In addition, figure 54 shows that use of shims up to the loading-platen angle of 1.2 degrees has no significant effect on the results of the large-specimen test. However, the loading-platen angle of 1.2 degrees for large-specimen testing is practically too large, and it is important to address the cause of the large angle to prevent any other potential issues. Therefore, the loading-platen angle of 0.6 degrees was selected as a conservative limit to establish the threshold for both the large- and small-specimen geometries.

Direct measurement of the loading-platen angle is difficult, but it is possible to measure the angle indirectly by checking the gap between the loading platen and the AMPT platen. This gap was ultimately suggested for incorporation into the revised standards. To establish the allowable gap, it was necessary for the research team to make some assumptions about the loading-platen geometry. Based on the current configuration of the AMPT, loading-platen diameters ranging from 80 to 130 mm were used, as shown in figure 55. In the ruggedness study, a loading platen with an 80-mm diameter was used for small-specimen testing, but a loading platen with a 130-mm diameter is considered feasible, as this is the diameter of the AMPT platen. In figure 55, d, g^e, and g^m represent the loading-platen angle, effective gap, and measured gap, respectively. The g^e parameter subscripts of 1 and 2 indicate the specimen edge and loading platen edge, respectively. The effective gap is the gap that affects strain development upon tightening the attachment bolts, and the measured gap will be determined by the loading platen's diameter. In this ruggedness study, the research team used a loading platen with an 80-mm diameter, and the effective gap was 0.6 mm. For loading platens with a 130-mm diameter, the team used the same effective gap of 0.6 mm to calculate the loading-platen angle and determined the measured gap by the loading-platen angle and diameter, as shown in figure 55. The allowable gap for a loading platen with an 80-mm diameter is 0.8-mm, and the gap for a loading platen with a 130-mm

diameter is 1.0-mm. For large-specimen testing, the research team considered only a loading platen with a 130-mm diameter. Consequently, the maximum allowable gap is 1 mm for large-specimen geometry and 0.8 mm for small-specimen geometry. The calculations are presented in the appendix. For smaller gaps, the research team proposes using shims to mitigate strain development when tightening the attachment bolts.
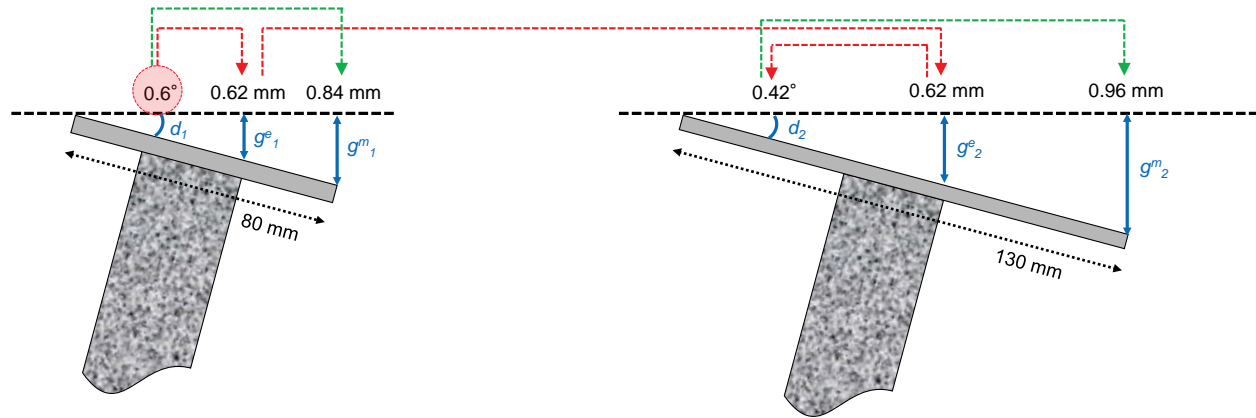
Source: FHWA.

Figure 55. Illustration. Schematics showing allowable gaps between loading platen and AMPT platen with possible loading-platen geometries.

ANALYTICAL TRANSFER-TIME STUDY

The research team conducted testing to calibrate a heat-transfer model to develop parameters for temperature-conditioning time using the allowable temperature-error threshold of ±0.5°C established from the results of the ruggedness evaluation. Subsequently, the team used the model to determine the effects of laboratory and handling conditions on specimen temperature and time to reequilibrate under different setup scenarios. The experiments consisted of first placing a specimen into an external chamber while monitoring the time needed for the surface and center of the specimen to reach temperature equilibrium. The test specimens were a surface mixture from North Carolina comprising 20 percent reclaimed asphalt pavement by aggregate mass. For the testing conditions, the specimens had an air-void content of 4 percent, a void-in-mineral-aggregate content of 15.8 percent, and a voids-filled-with-asphalt value of 75 percent. One thermocouple that was insulated from direct contact with the surrounding air was attached to the surface of the specimen, and another was inserted into the middle of the specimen. The specimen was then removed from the chamber and placed onto the bottom loading plate of the AMPT machine. The specimen rested at this condition for approximately 20 min while the air temperature, specimen center, and specimen surface temperatures were recorded. The AMPT testing cell was then closed and set to 18°C, and the specimen temperatures monitored for approximately 2 h (specimen 1) or 6.5 h (specimen 2); these times were selected conservatively to ensure thermal equilibration was achieved. The purpose of these experiments was to measure temperature change under conditions similar to those that a specimen would experience while undergoing transfer, setup, and temperature reequilibrium in the AMPT testing cell as well as to calibrate the thermodynamic convection constant of a typical laboratory and the AMPT instrument itself. The standard heat-flow function for an axisymmetric rod with uniform temperature along its length, as shown in figure 56, was implemented into a MATLAB®

script.[26] Boundary conditions were assumed to be of a Neumann type and governed by convective flow, as shown in figure 57. Specimens of 100-mm diameter by 130-mm tall and 38-mm diameter by 110-mm tall were simulated.

$$r\rho C_h \frac{\partial T_{x,t}}{\partial t} = \frac{\partial}{\partial r}\left(k_c r \frac{\partial T_{x,t}}{\partial r} \right)$$

Figure 56. Equation. Standard heat-flow function.

Where:
  r = radius (m).
  ρ = density (kg/m^3).
  C_h = specific heat (W-s/kg-°C).
  $T_{x,t}$ = current temperature at position x and time t (°C).
  k_c = thermal conductivity (W/m-°C).

$$\lambda r\left(T_{ext} - T_{x,t} \right) = \vec{n}\left(k\nabla T_{x,t} \right)$$

Figure 57. Equation. Neumann-type boundary condition governed by convective flow.

Where:
  λ = convective heat coefficient (W/m^2/°C).
  $\vec{n}$ = normal unit vector.
  T_{ext} = external temperature (°C).

The research team calibrated the convective temperature constants using the measured air temperature as input and used a trial-and-error approach to matching the experimental data. This calibration yielded a laboratory convective flow coefficient, λ_{lab}, of 15 W/m^2/°C and an AMPT convective coefficient, λ_{AMPT}, of 20 W/m^2/°C for the AMPT instrument. The thermal conductivity, k_c, was 2 W/m-°C, the specific heat was 900 W-s/kg-°C, and the density, ρ, was 2320 kg/m^3. The thermal coefficients were taken as representative based on the values reported elsewhere, while the density was the actual density of the test specimens. [27] The results of the calibration process are shown in figure 58 through figure 61.
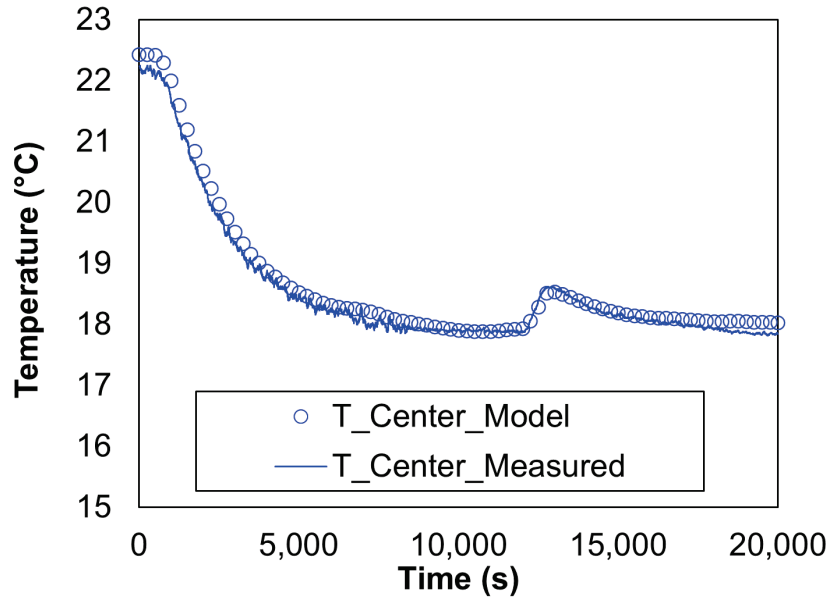
Figure 58. Graph. Predicted and measured temperatures at the at the specimen 1 center.
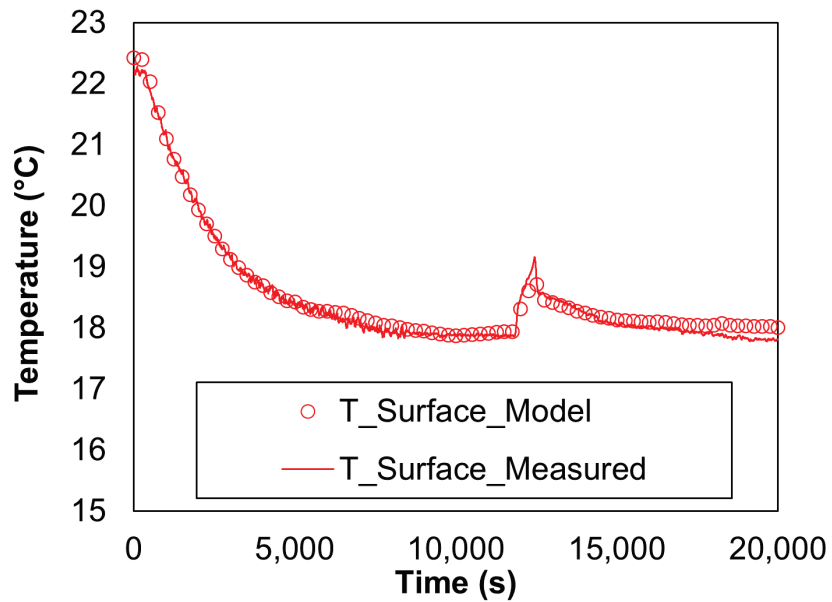
Figure 59. Graph. Predicted and measured temperatures at the specimen 1 edge.
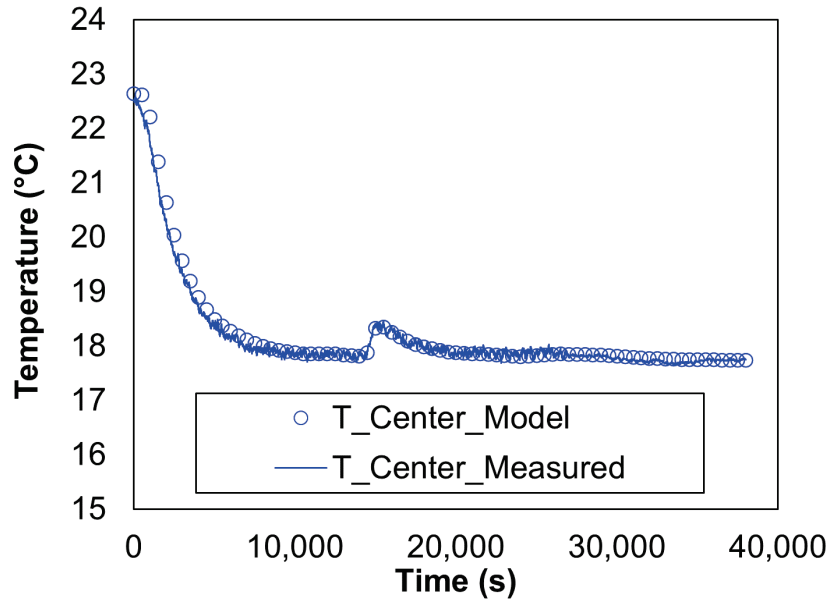
Source: FHWA.

Figure 60. Graph. Predicted and measured temperatures at the specimen 2 center.
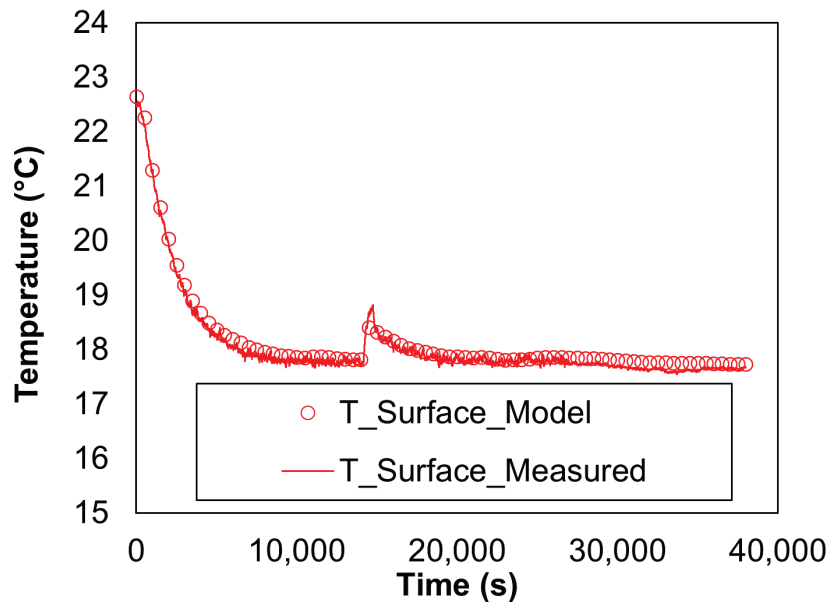
Source: FHWA.

Figure 61. Graph. Predicted and measured temperatures at the specimen 2 edge.

The research team then used the model to simulate the temperature response of an asphalt mixture under two different use conditions. In the first, termed AMPT conditioning, the research team predicted the time to reach equilibrium for a room-temperature specimen conditioned fully in the AMPT instrument. In the second, researchers assumed a three-step process: a specimen was first conditioned in an external chamber, then the specimen was transferred to the AMPT and connected to the machine, and finally the testing cell was closed and returned to temperature. The total time for these three steps was simulated to be between 2.5 and 20 min. Based on the

61

measurements using the AMPT machine, the testing cell was assumed to cool at a rate of 2°C per min. Simulations were carried out in both scenarios for room temperature-to-testing temperature differences between 1°C and 15°C. Note that, based on the assumed cooling rate, some working scenarios were not physically possible for the larger temperature differences.

The research team assumed sufficient equilibrium was achieved when the temperature at the center of the specimen was within ±0.5°C of the target temperature. This range was used even though the ruggedness study suggested a range of ±2°C would be acceptable to account for the tolerance of the testing equipment and the external temperature chamber, which are both ±0.5°C. This temperature precision is common to other testing standards and likely represents a mechanical limitation of common refrigerant-based cooling systems. As a result, however, two labs using standard equipment could already have a 2°C difference in temperatures from the point when the specimens are removed from the external chamber. These differences would theoretically suggest that sufficient equilibrium was achieved when the center of the specimen was within ±1.0°C of the target. However, the ±0.5°C limit was selected after recognizing that the difference in temperature between the fingerprint portion of the test and the beginning of the fatigue test was also important. The fingerprint test is used to define the *DMR*, which accounts for sample-to-sample variability and helps to reduce overall scatter in the test results. This test is performed at a minimum of 5 min prior to the fatigue test, and any differences in temperature that occur between the fingerprint and the beginning of the fatigue test will result in overall reduced repeatability and reproducibility of the test results. Underwood showed that the modulus of asphalt mixture changes by a power of approximately 2.5 in terms of temperature increases; that is, if the temperature increases from 10°C to 20°C, the modulus would decrease by a factor of 5.7 ($2^{-2.5}$).[28] Using this relationship, the thermal equilibrium model suggests that, when a threshold of ±0.5°C is used, the additional 5 min of equilibrium during fingerprint testing translates to an average error in *DMR* of between 0.6 and 2.3 percent, whereas the 1°C threshold translates into an average error of between 0.4 and 4.2 percent. Further, the maximum error in the case of the 1°C threshold (for the cases evaluated) was 9.4 percent.

The results of the analysis with the ±0.5°C limit are summarized in table 20 and table 21. Note that some temperatures require no reequilibrium time. Also note that the working time includes time to return the testing chamber back to the testing temperature. This process can take longer than 5 min in some cases, and these instances are marked with a dash in table 20 and table 21.

Table 20. Summary of reequilibrium time as a function of room temperature to test temperature difference and working time for large specimen (min).

ΔT (Temperature of Room, °C)*	Direct Condition	Less than 5 min of Working + Equilibrium Time	5-10 min of Working + Equilibrium Time	10-20 min of Working + Equilibrium Time	More than 20 min of Working + Equilibrium Time
15	140	—	80	90	140
14	130	—	80	90	130
13	130	—	80	90	130
12	130	—	70	80	130
11	120	—	70	80	120
10	120	—	70	80	120
9	120	50	60	70	120
8	110	40	60	70	110
7	110	30	50	60	110
6	100	30	40	60	100
5	100	20	40	50	100
4	90	0	30	40	90
3	80	0	20	30	80
2	60	0	0	0	60
1	40	0	0	0	40

*Test temperature.
—Not applicable.

Table 21. Summary of reequilibrium time as a function of room temperature to test temperature difference and working time for small specimen (min).

ΔT (Temperature of Room, °C)	Direct Condition	Less than 5 min of Working + Equilibrium Time	5-10 min of Working + Equilibrium Time	10-20 min of Working + Equilibrium Time	More than 20 min of Working + Equilibrium Time
15	60	—	40	50	60
14	60	—	40	50	60
13	60	—	40	50	60
12	50	—	40	40	50
11	50	—	40	40	50
10	50	—	40	40	50
9	50	30	40	40	50
8	50	30	30	40	50
7	50	30	30	40	50
6	40	20	30	30	40
5	40	20	30	30	40
4	40	20	20	30	40
3	30	10	20	20	30
2	30	0	10	20	30
1	20	0	0	0	20

*Test temperature.
—Not applicable.

PROPOSED CONTROLS FOR THE EXPERIMENTAL FACTORS

Based on the results shown in table 18 and table 19, combined with the results of the supplementary studies of end-failure mitigation and transfer time, the research team established conservative controls for the experimental factors, as shown in table 22.

The team proposes a specimen air-void tolerance of ±0.5 percent from the target because tighter control of air voids is not practically feasible.[29] Lab B ruggedness-evaluation results for small specimens suggest that a tolerance of ±0.4 percent is required to limit effects on FlexPAVE percentage-cracking predictions. In addition, the air-void content for each ruggedness level has ±0.5 percent tolerance range. In other words, the research team assumed that a target air-void content of ±0.5 percent can represent the specific target air-void content; therefore, the proposed tolerance for air-void content cannot be narrower than the range of air-void content ±0.5 percent. Based on the statistical analysis results for air-void content, tolerance for large-specimen geometry could be a target air-void content of ±0.8 percent. However, it is important to establish the same tolerance range for both geometries because a mixture can be used to conduct cyclic fatigue tests in accordance with both standards for large and small specimens.

Table 18 and table 19 demonstrate that specimen height and rest period are insignificant in all cases; consequently, the proposed specimen-height tolerances are identical to the ruggedness levels. A larger tolerance range for specimen height is not possible given the actuator travel distance and platen thickness in at least one AMPT currently manufactured and in use within the United States. Rest period had no significant effect on the FlexPAVE results; therefore, a minimum limit of 5 min is proposed with no upper limit. The calculated tolerance ranges in table 18 and table 19 are used to guide refinement of the tolerance ranges for differences in diameter between the loading platen and large and small specimens given the specimen-diameter tolerances in AASHTO R 83-17 and AASHTO PP 99-19, respectively.[9,10]

Maximum limits for the gap between the loading platen and AMPT are based on the results of the study on end-failure mitigation, which demonstrate that when shims are used to limit strain development upon attaching the specimen to the AMPT, a loading-platen angle of up to 1.2 degrees can be used successfully. The maximum allowable gap limits presented in table 22 are based on possible loading-platen diameters and a conservative angle of 0.6 degrees.

Table 20 and table 21 present parameters for temperature-conditioning time established in the analytical transfer-time study using a heat-transfer analysis. This analysis yielded the equilibration time required for the specimen's temperature to reach within ±0.5°C of the target. As explained in the Analytical Transfer-Time Study section, the research team selected the ±0.5°C threshold based on a combination of factors, which included the calculated tolerance range for temperatures shown in table 18 and table 19, potential temperature errors introduced given the temperature tolerances of testing equipment and external temperature chambers, and the possible change in test-specimen temperature during the rest period between fingerprint loading and subsequent fatigue loading.

Table 22. Proposed controls for the experimental factors.

Factor	Large Specimen	Small Specimen
Air void	Target air void ±0.5%	Target air void ±0.5%
Height	127.5 to 132.5 mm	107.5 to 112.5 mm
Rest period	Minimum of 5 min	Minimum of 5 min
Platen diameter	100 to 105 mm	37.8 to 39.2 mm
Gap between loading platen and AMPT platen	Maximum 1 mm	Maximum 0.8 mm
Temperature	Target temperature ±0.5°C	Target temperature ±0.5°C

REFINEMENT OF *DMR* LIMITS BASED ON THE RUGGEDNESS-EVALUATION RESULTS

In the ruggedness study, three experimental ruggedness factors potentially affect *DMR*: air-void content, temperature, and loading-platen parallelism. *DMR* decreases as air-void content increases because fingerprint /*E**/ decreases when air-void content increases. Ruggedness experiments were conducted at a single test temperature; therefore, fingerprint /*E**/ was not affected. The study team conducted an analysis of the cyclic fatigue testing using two different temperature levels, however, which were determined by the ruggedness experimental design. Consequently, *DMR* was affected by the /*E**/ at the temperature level and frequency from the mastercurve. *DMR* increases as the ruggedness analysis temperature increases. Another factor, which may affect *DMR* in the ruggedness study, is loading-platen parallelism. When the loading platen has a nonparallel condition, the test specimen can be damaged during the bolt-tightening procedure, which causes the fingerprint /*E**/ to decrease.

Table 23 presents the ruggedness experimental plan for air void, temperature, and loading-platen parallelism. Based on the effect of air void and temperature on the *DMR* value, determinations 2 and 3 should have lower *DMR* values, and determinations 6 and 7 should have higher *DMR* values than those for other determinations. In addition, determinations 2, 4, 6, and 8 should not be considered in the linear regression analysis due to the nonparallel condition issue on the *DMR* calculation, with the exception of 19-mm NMAS mixture ruggedness testing. In testing the ruggedness of the 19-mm mixture, shims were used in both parallel and nonparallel conditions; therefore, unlike other mixtures, all determinations resulting from ruggedness testing using the 19-mm mixture can be considered in the linear regression analysis. For other mixtures, determinations 1, 3, 5, and 7 were considered in the linear regression analysis. The average *DMR* values used in the linear regression analysis are presented in figure 62 through figure 67, and the error bars indicate the standard deviations. Determinations 3 and 7 for all 9.5-mm and 25-mm mixtures have the lowest and highest *DMR* values as expected. For the 19-mm mixture, determinations 2 and 3 have lower *DMR* values, and determinations 6 and 7 have higher *DMR* values than those for other determinations.

Table 23. Ruggedness experimental plan for air void and temperature.

Factor	Det. 1	Det. 2	Det. 3	Det. 4	Det. 5	Det. 6	Det. 7	Det. 8
Air void	High	High	High	High	Low	Low	Low	Low
Temperature	High	Low	Low	High	Low	High	High	Low
Loading-platen parallelism	Parallel	Nonparallel	Parallel	Nonparallel	Parallel	Nonparallel	Parallel	Nonparallel

Det. = determination.
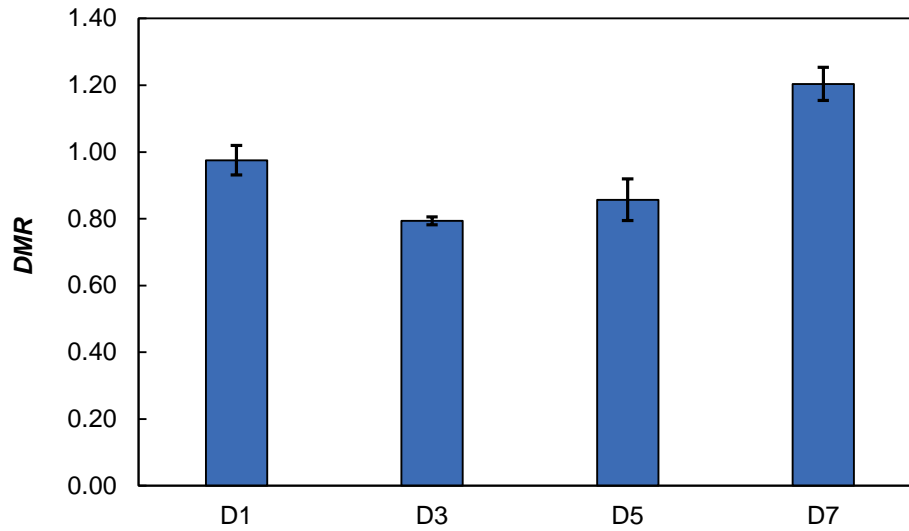
Source: FHWA.
D = determination.

Figure 62. Graph. Average *DMR* values for lab A large-specimen testing using 9.5-mm NMAS mixture.
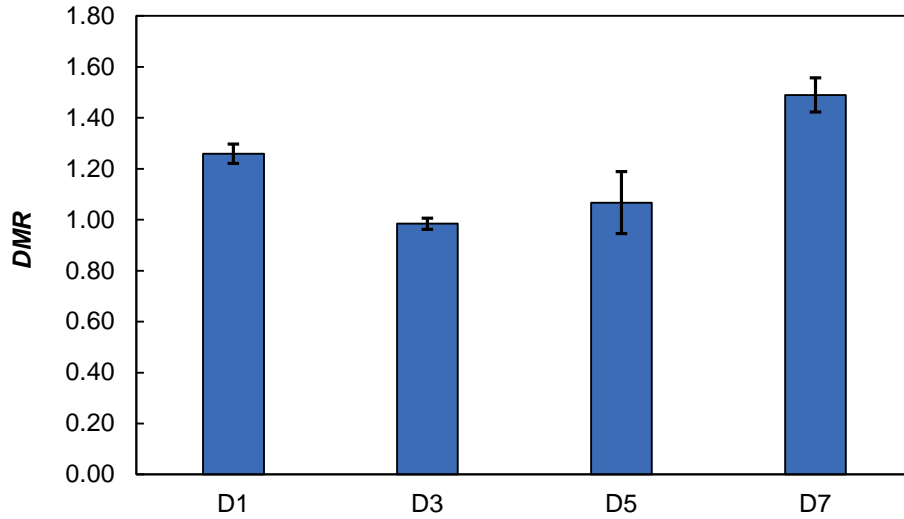
Figure 63. Graph. Average *DMR* values for lab B large-specimen testing using 9.5-mm NMAS mixture.

Figure 64. Graph. Average *DMR* values for large-specimen testing using 25-mm NMAS mixture.

Source: FHWA.
D = determination.

Figure 65. Graph. Average *DMR* values for lab A small-specimen testing using 9.5-mm NMAS mixture.
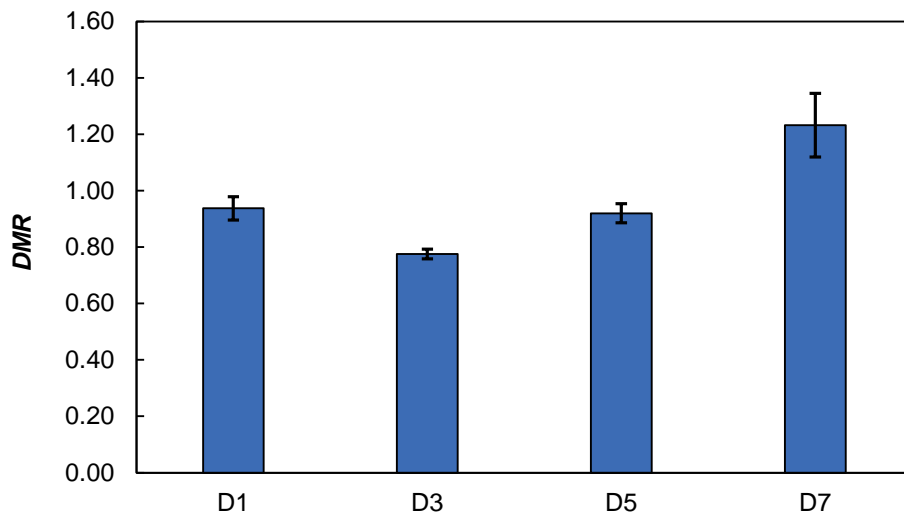
Source: FHWA.
D = determination.

Figure 66. Graph. Average *DMR* values for lab B small-specimen testing using 9.5-mm NMAS mixture.
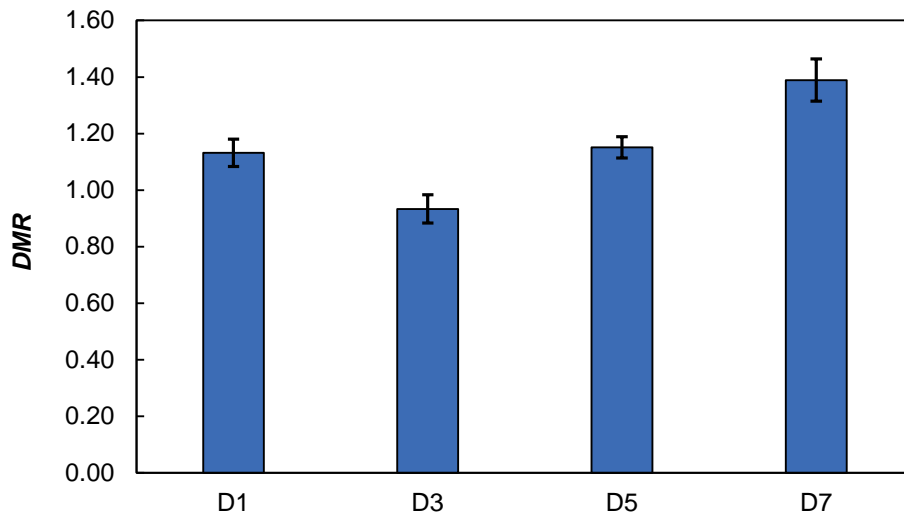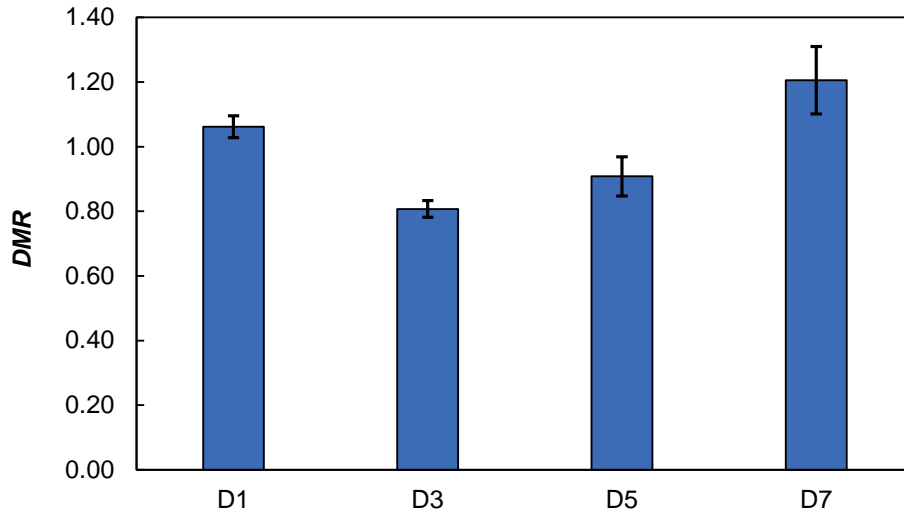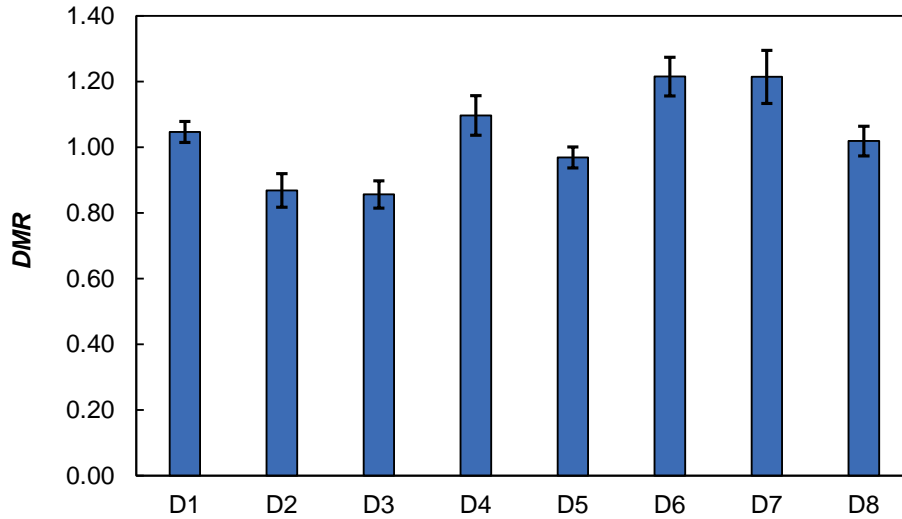
Source: FHWA.
D = determination.

Figure 67. Graph. Average *DMR* values for small-specimen testing using 19-mm NMAS mixture.

To establish the *DMR* limits based on the ruggedness-evaluation results, the research team conducted linear regression analyses using two ruggedness experimental factors: air void and temperature. To evaluate the ruggedness of the 9.5-mm and 25-mm NMAS mixtures, the team used *DMR* values from determinations 1, 3, 5, and 7 in the linear regression analyses. The team also used *DMR* values from all determinations for 19-mm NMAS mixture. Figure 68 to figure 73 show linear regression analysis results for each ruggedness evaluation, with the predicted and measured *DMR* values generally indicating good agreement. The main effects of air void and temperature on *DMR* for each ruggedness mixture are presented in figure 74.
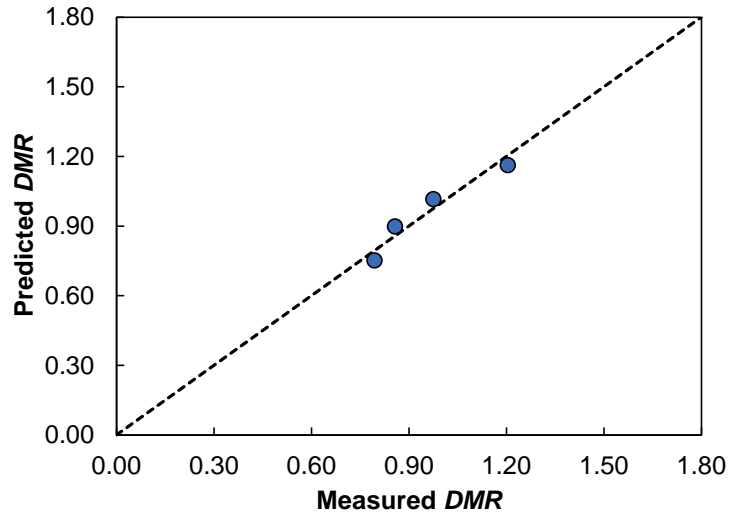
Figure 68. Graph. Predicted versus measured *DMR* values for lab A large-specimen testing using 9.5-mm NMAS mixture.
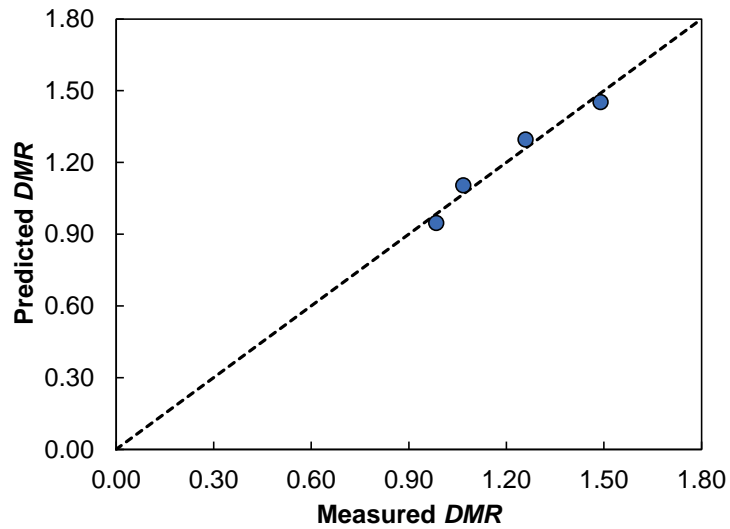
Figure 69. Graph. Predicted versus measured *DMR* values for lab B large-specimen testing using 9.5-mm NMAS mixture.

Figure 70. Graph. Predicted versus measured *DMR* values for large-specimen testing using 25-mm NMAS mixture.

Figure 71. Graph. Predicted versus measured *DMR* values for lab A small-specimen testing using 9.5-mm NMAS mixture.

Figure 72. Graph. Predicted versus measured *DMR* values for lab B small-specimen testing using 9.5-mm NMAS mixture.

Figure 73. Graph. Predicted versus measured *DMR* values for small-specimen testing using 19-mm NMAS mixture.
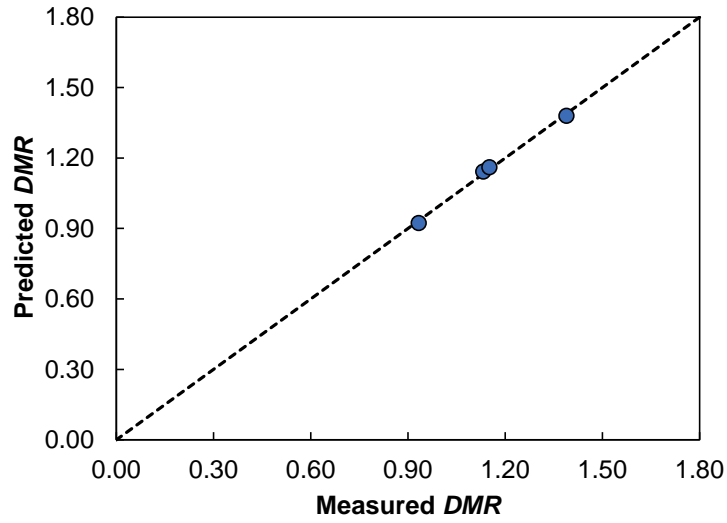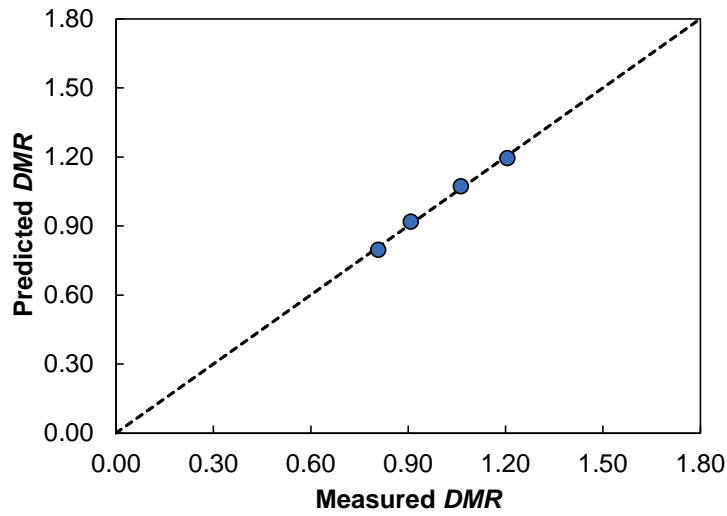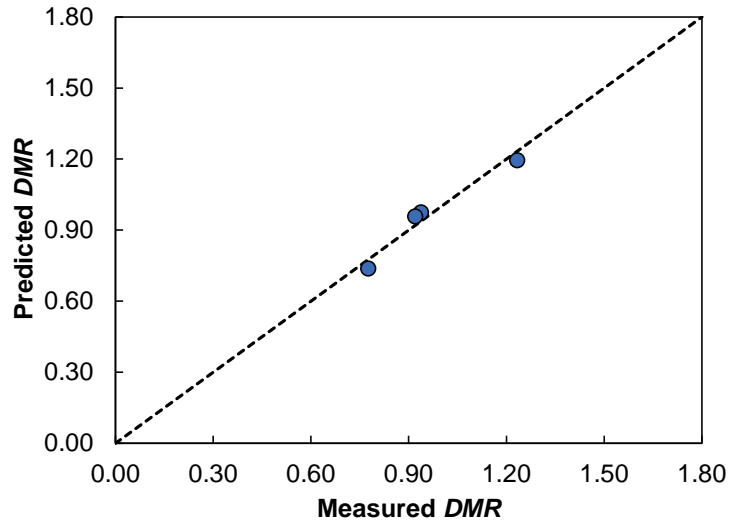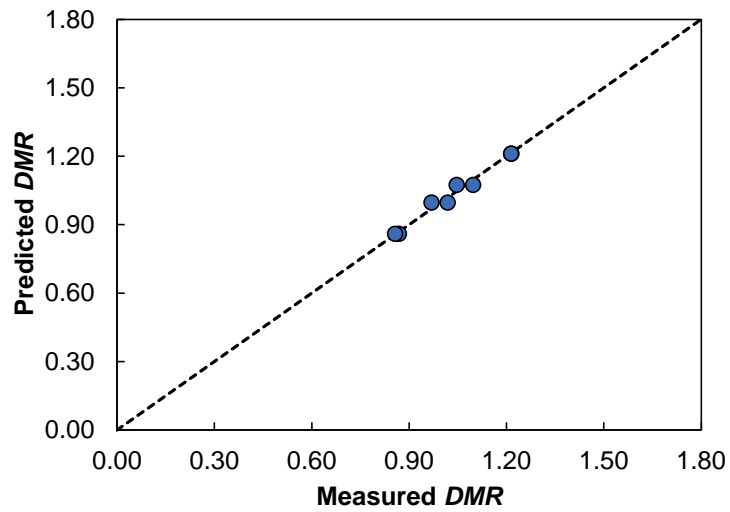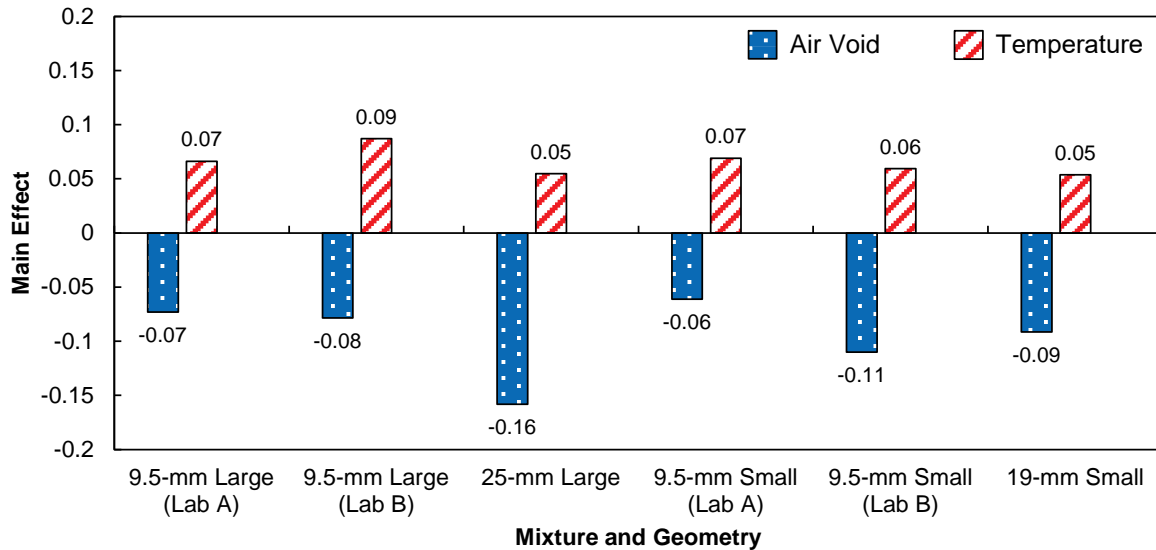
Source: FHWA.

Figure 74. Graph. Main effects of air void and temperature on *DMR* for different ruggedness-evaluation conditions.

In this ruggedness study, target air-void content of ±0.5 percent was proposed based on the statistical analysis. Regarding the temperature effect, although the ruggedness study suggested a variance of ±2°C would be acceptable, two labs using standard equipment could already have a 2°C difference in temperatures from the point when the specimens are removed from the external chamber, as discussed within the Analytical Transfer-Time Study section. These differences would theoretically suggest that sufficient equilibrium was achieved when the center of the specimen was within ±1.0°C of the target. Therefore, the research team used a target air void of ±0.5 percent and a target temperature of ±1.0°C to establish the *DMR* limits. The calculated allowable *DMR* limits based on the air void and temperature effects were presented in table 24. The ruggedness-evaluation results for the large specimen of the 25-mm NMAS mixture have the widest allowable *DMR* range, which is ±13.4 percent. Thus, the maximum *DMR* limit of ±15 percent (after rounding to the nearest 5 percent) is proposed from this study. This threshold will be further evaluated through the ILS.

Table 24. Calculated *DMR* limits based on the main effects.

Mixture (Lab)	9.5-mm Large (Lab A)	9.5-mm Large (Lab B)	25-mm Large	9.5-mm Small (Lab A)	9.5-mm Small (Lab B)	19-mm Small
Calculated *DMR* limits	±10.3%	±12.6%	±13.4%	±10.0%	±11.4%	±10.0%

ESTABLISHING A DATA-QUALITY INDICATOR TO ENSURE PROPER PID SETTINGS

The research team reviewed existing data from the five initial cycles of cyclic fatigue testing conducted as part of the ruggedness evaluation as well as the data acquired in the FHWA PRS

73

shadow projects to establish limits to ensure proper PID settings. The PRS shadow-project data encompassed six mixtures from four laboratories.

The PID is a control-loop mechanism employing feedback that is widely used in control systems. The mathematical form of the PID control function is shown in figure 75.[29]

$$u(t) = K_p e(t) + K_i \int_0^t e(t')dt' + K_d \frac{de(t)}{dt}$$

Figure 75. Equation. PID control function.[29]

Where:
   $u(t)$ = output.
   $e(t)$ = error.
   K_p = proportional gain.
   K_i = integral gain.
   K_d = derivative gain.

Erroneous PID settings can lead to differences between the input and resultant strain level as well as nonsinusoidal loading. AASHTO TP 107-18 and AASHTO TP 133-19 provide guidance for selecting input strain levels based on the $|E^*|$ at the test condition.[1,2] If the input strain and resultant onspecimen strain levels differ, the strain selection guide cannot be used to estimate the appropriate input strain to yield a proper fatigue life. Furthermore, sinusoidal loading is required for interpretation of the test results using the S-VECD model.

Since the cyclic fatigue test is a displacement-controlled actuator test, the research team used standard error of actuator strain to ensure proper PID settings. This value is calculated as the error between the actuator command and the actual measured data, as shown in figure 76. The actuator command, which starts at time zero, can be generated based on least-squares fit of a sinusoid using given amplitude and frequency.[30] However, the measured data need to be shifted to calculate the standard error correctly because there is a small time lag at the beginning of the test in the output data. Due to signal noise, it is difficult to find the starting point of the sinusoid. However, instead of the starting point of the sinusoid, the peak point in each cycle of the sinusoid can be used to determine the time lag. The peak points should be at 0.05, 0.15, 0.25, 0.35, and 0.45 s because the frequency is 10 Hz, and the time lag at the peak points can be determined from the measured data. Based on the average time lag of the initial five cycles, the data can be shifted to remove the time lag. After the measured data are shifted, the actuator strain standard error for each cycle is calculated as shown in figure 76, and then the average standard error is calculated using each standard error in cycles two through five. Note that the average standard error could be calculated using the data point from cycles two through five all at once, but the average error is slightly lower than the average error from individual calculations for each cycle. Although the difference between the average error was small (about 0.1 to 0.2 percent), the average error from the individual calculation for each cycle was used as a conservative limit in this study.

$$SE = \sqrt{\dfrac{\displaystyle\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n-4}} \left(\dfrac{100\%}{\hat{x}_o} \right)$$

**Figure 76. Equation. Standard error of actuator strain.**

Where:

SE = standard error of actuator strain.

x_i = measured actuator strain at point i.

$\hat{x}_i$ = predicted actuator strain at point i from the sinusoid.

n = total number of data points collected during testing.

$\hat{x}_0$ = amplitude of the sinusoid.

Figure 77 shows the example results comparing the command and measured actuator strain in the initial five cycles of an AMPT cyclic fatigue test when the PID settings were properly tuned. As shown, the command and measured actuator strains are in good agreement, and in this case, the average standard error of the actuator strain in the five initial cycles is 2.5 percent. Figure 78 and figure 79 provide example results for actuator strain in the five initial cycles of the AMPT cyclic fatigue tests when the PID settings were erroneous. In the example shown in figure 78, the deviation between the measured and actuator-command strains is consistent throughout the five cycles, with an average error for actuator strain of approximately 12 percent. The example in figure 79 shows a case where the first cycle has a relatively large discrepancy between the command and measured actuator strains, but this error is considerably less in subsequent cycles. The standard error of the actuator strain in the first cycle is approximately 16 percent, and the average error in cycles two through five is approximately 8 percent. The effect of the overshoot in the actuator strain in the first cycle was investigated by comparing the damage-characteristic curve and D^R results from this test with other tests that had proper PID settings. The damage-characteristic curves and D^R failure criterion results from those experiments were in good agreement, and there were no clear effects of erroneous PID setting that resulted in the overshoot on damage-characteristic curve and D^R failure criterion results from the first cycle, as shown in figure 80 and figure 81, respectively. Therefore, the research team proposes that the data-quality indicator to ensure proper PID setting be based on the average standard error in cycles two through five of the cyclic fatigue tests.
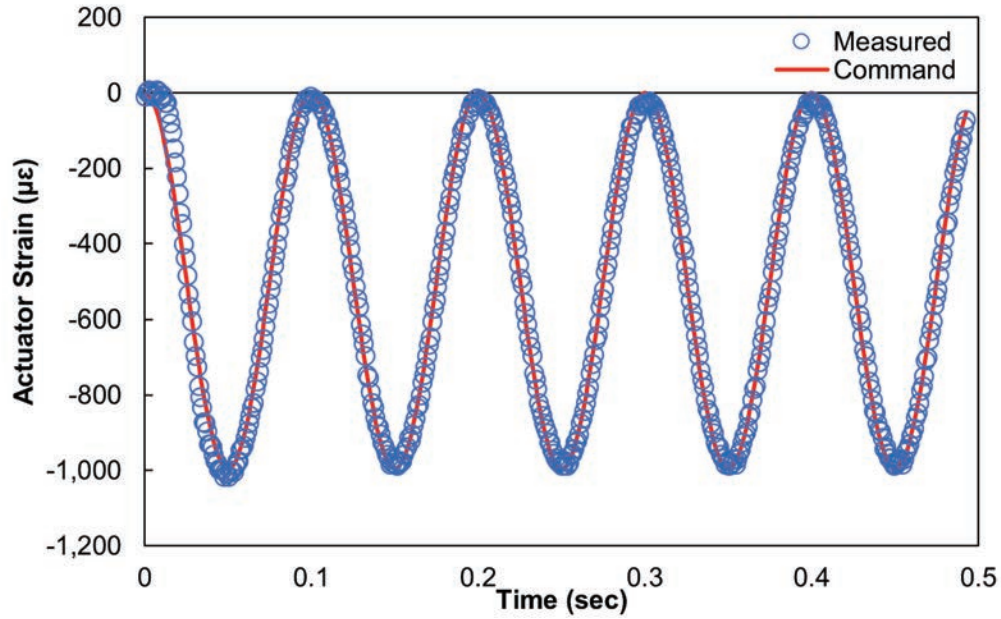
Figure 77. Graph. An example of actuator strain for five initial cycles with proper PID tuning.
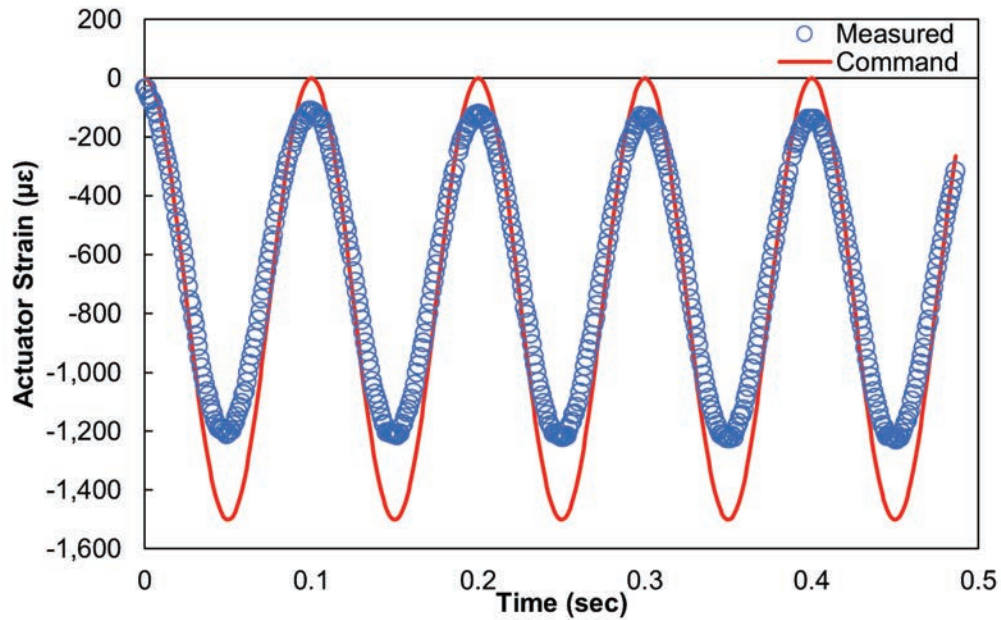
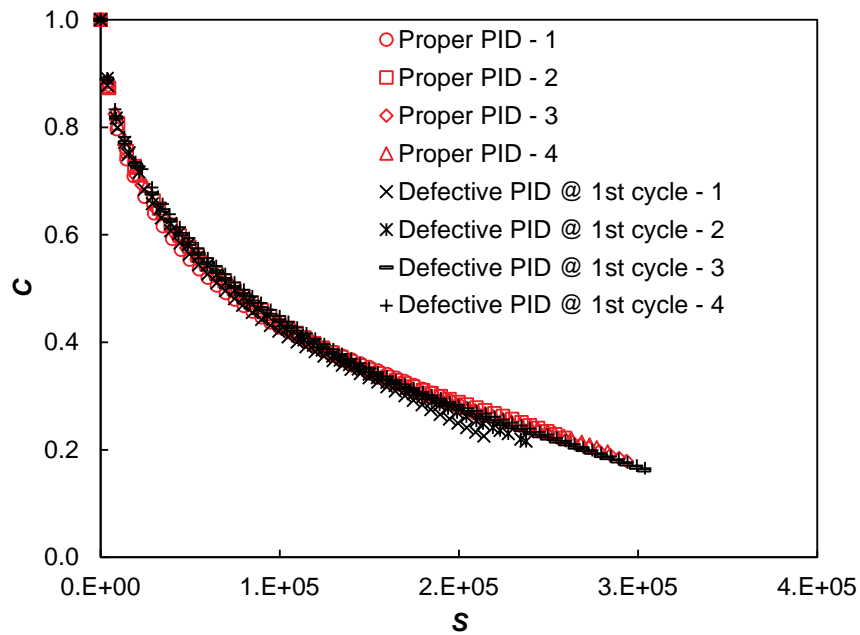**Figure 78. Graph. An example of actuator strain for the initial five cycles with defective PID tuning.**
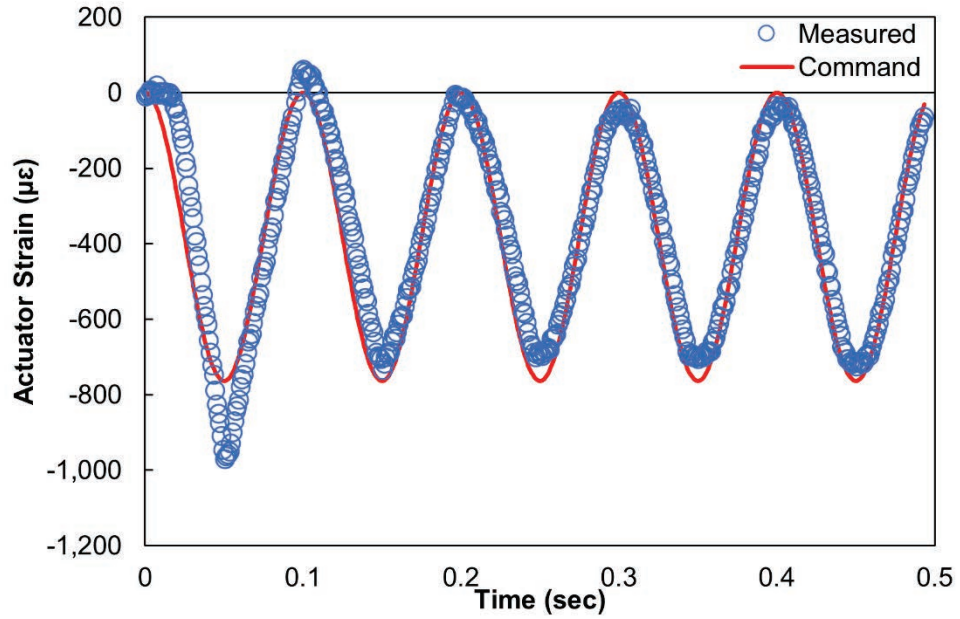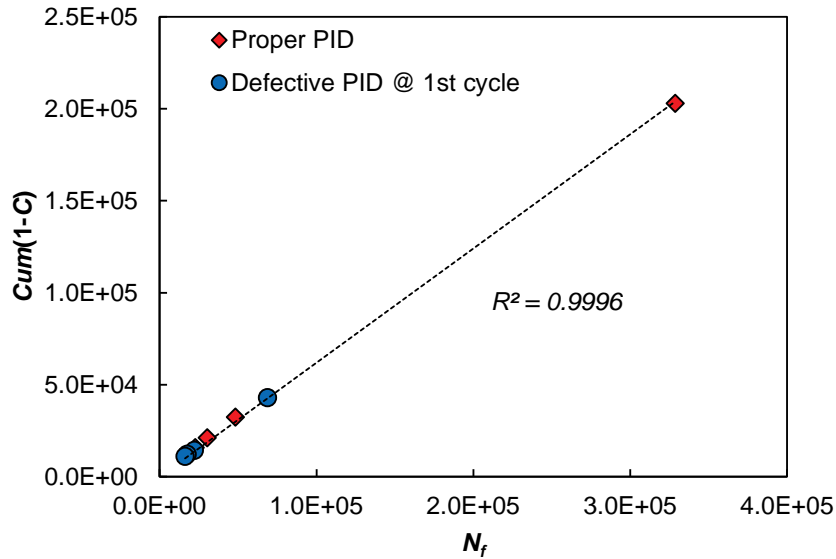
Source: FHWA.

Figure 79. Graph. An example of actuator strain for the initial five cycles with defective PID tuning at the first cycle.

Source: FHWA.

Figure 80. Graph. Comparison of damage-characteristic curves between proper PID tuning and defective PID tuning at the first cycle.

Source: FHWA.

Figure 81. Graph. Comparison of D^R failure criteria between proper PID tuning and defective PID tuning at the first cycle.

The research team combined the measured errors for actuator strain from the results of the ruggedness evaluation with data acquired in the FHWA PRS shadow projects to establish a threshold for the proposed data-quality indicator. Figure 82 shows the average standard error of actuator strain in cycles two through five for each test specimen from the six shadow-project mixtures mentioned earlier. Table 25 shows the general mixture information. Figure 82 demonstrates that in most tests, the average standard error of actuator strain in cycles two through five is less than 10 percent. However, mixtures C and F both include several tests with average errors that exceed 10 percent. Investigation of the damage-characteristic curves for mixtures C, E, and F demonstrated that 10 percent is the maximum average standard error of actuator strain that can be tolerated without impacting the test results. Figure 83 through figure 88 show the damage-characteristic curves for mixtures A through F, respectively. The numbers in parentheses in the figure legends indicate the average standard error of actuator strain in cycles two through five. The results in figure 85 and figure 88 clearly show that the damage-characteristic curves with actuator strain standard errors exceeding 10 percent differ from the curves derived from test results where the standard error was lower. Figure 87 shows the damage-characteristic curves for mixture E. Again, the numbers in parentheses in the figure legend indicate the average standard error of actuator strain in cycles two through five. Figure 87 shows no clear bias in the results of the test, with the average standard error of actuator strain of 9 percent compared to the results of tests with lower errors. Thus, the research team proposes a maximum limit of 10 percent on the average standard error of actuator strain in cycles two through five to ensure proper PID settings. This threshold will be further evaluated through the ILS.

Table 25. Shadow project mixture information.

Mixture	NMAS (mm)	Binder PG	RAP content (%)
Mixture A	12.5	PG 64E-28	20
Mixture B	12.5	PG 64E-28	20
Mixture C	12.5	PG 64-28	20
Mixture D	12.5	PG 64-28	20
Mixture E	9.5	PG 64-22	30
Mixture F	12.5	PG 64-22	20
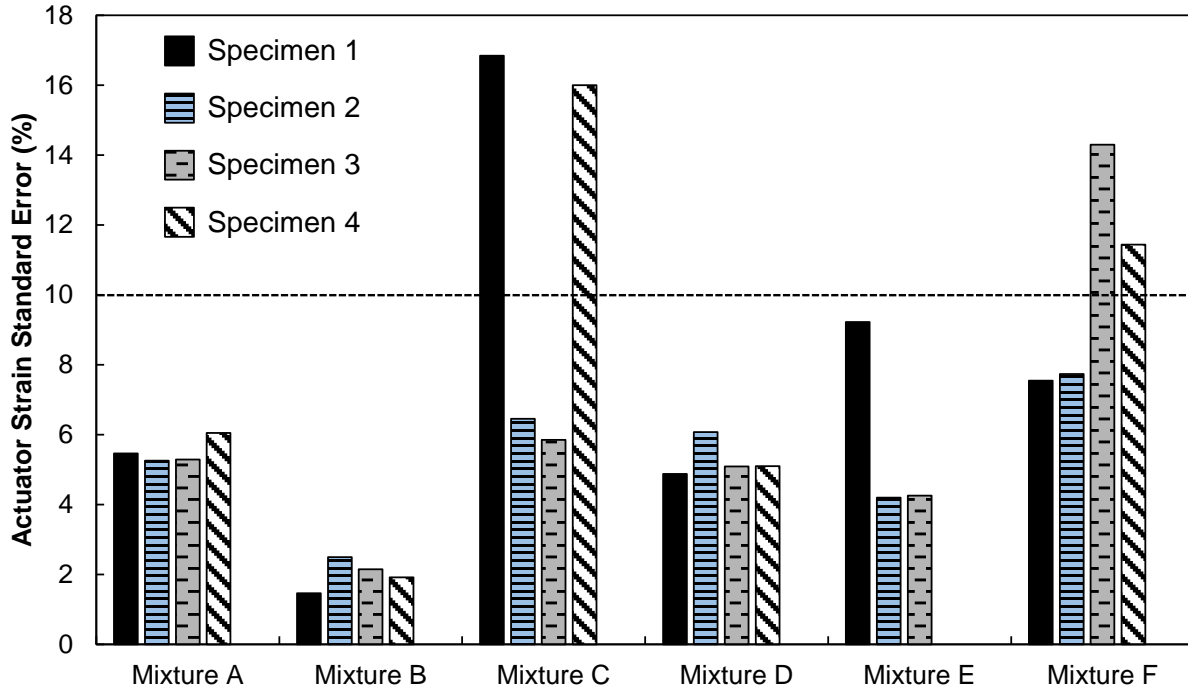
Source: FHWA.

Figure 82. Graph. Actuator strain standard errors for FHWA PRS shadow projects.
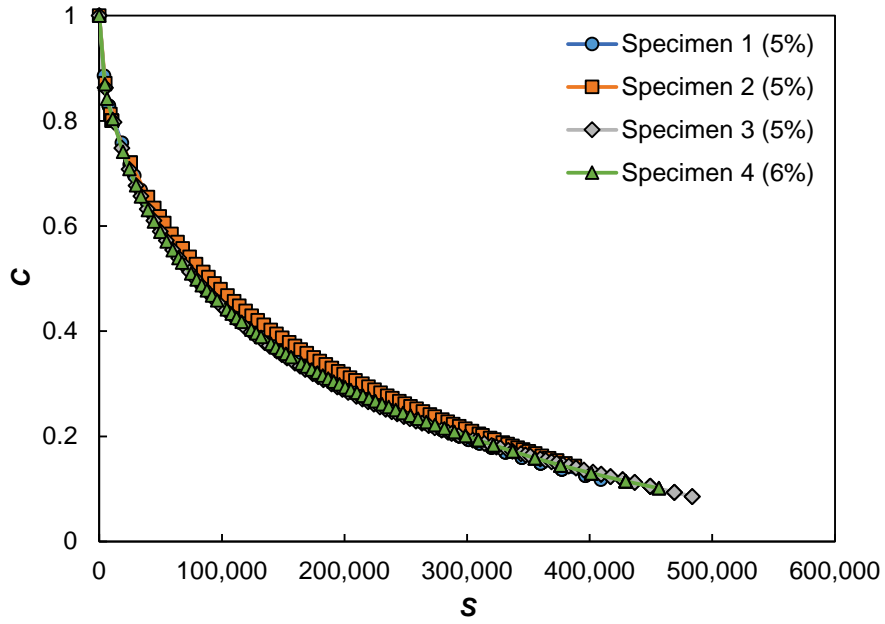
Note: The parenthetical numbers in the legend indicate the average standard error for actuator strain in cycles two through five.

Figure 83. Graph. Damage-characteristic curves for mixture A
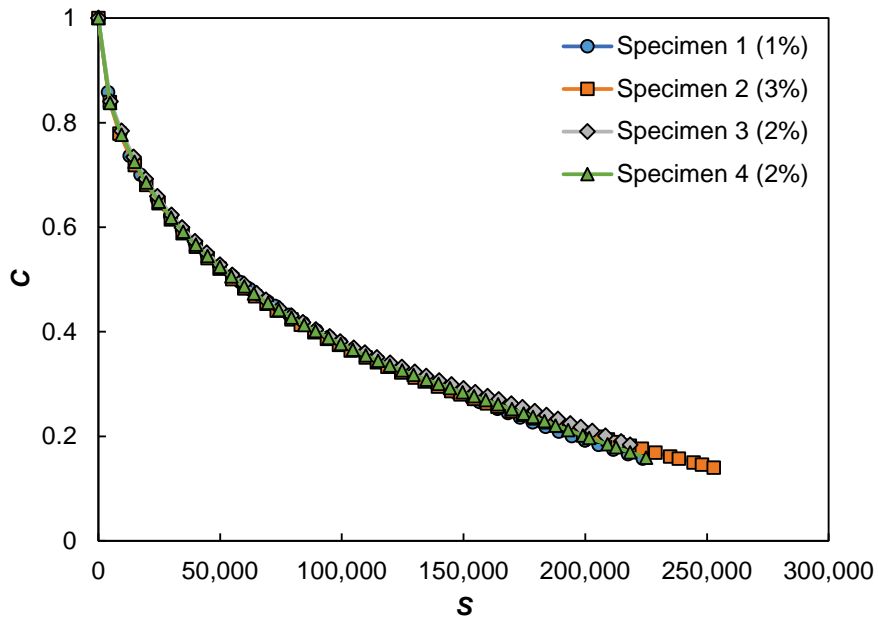
Note: The parenthetical numbers in the legend indicate the average standard error for actuator strain in cycles two through five.

Figure 84. Graph. Damage-characteristic curves for mixture B.
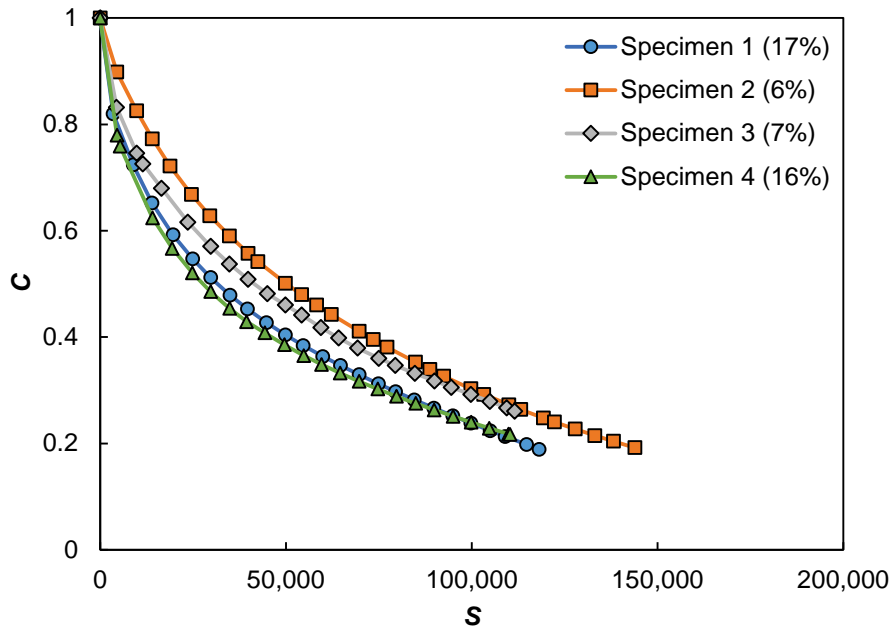
Source: FHWA.

Note: The parenthetical numbers in the legend indicate the average standard error for actuator strain in cycles two through five.

Figure 85. Graph. Damage-characteristic curves for mixture C.
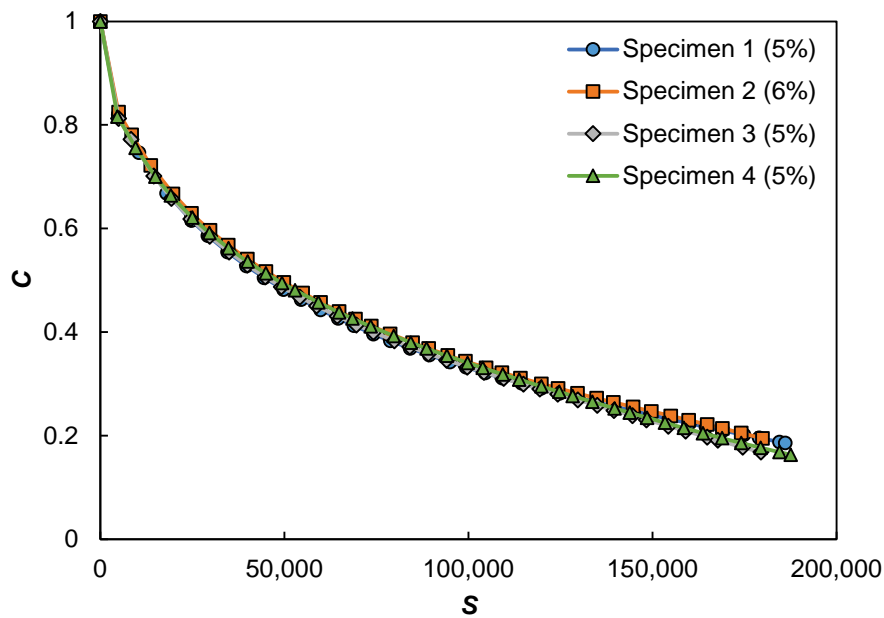
Source: FHWA.

Note: The parenthetical numbers in the legend indicate the average standard error for actuator strain in cycles two through five.

Figure 86. Graph. Damage-characteristic curves for mixture D.
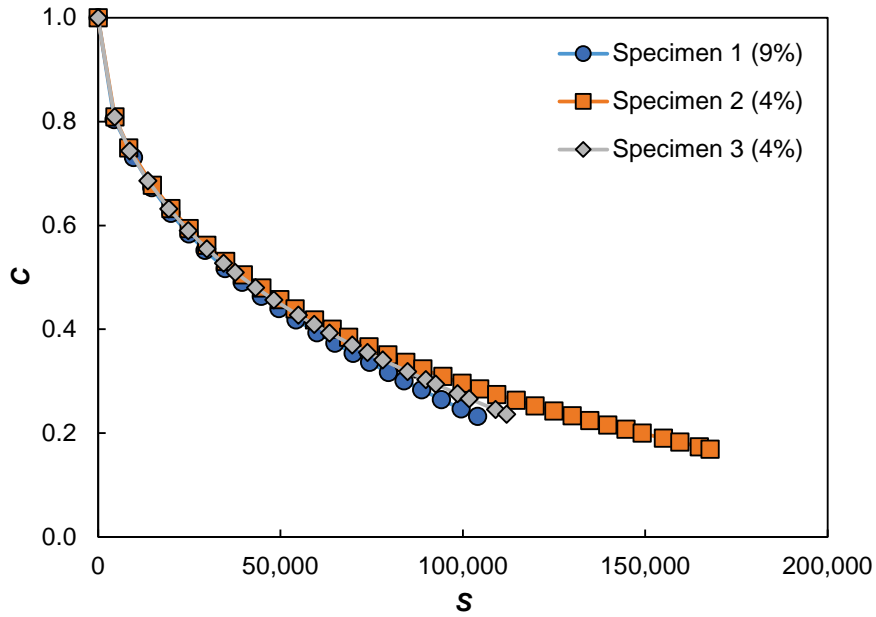
Source: FHWA.
Note: The parenthetical numbers in the legend indicate the average standard error for actuator strain in cycles two through five.

Figure 87. Graph. Damage-characteristic curves for mixture E.
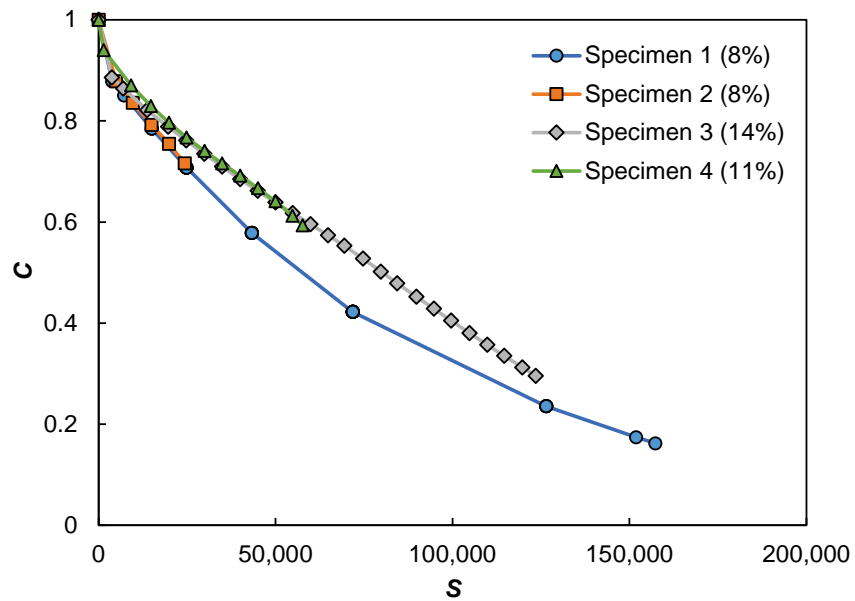
Source: FHWA.
Note: The parenthetical numbers in the legend indicate the average standard error for actuator strain in cycles two through five.

Figure 88. Graph. Damage-characteristic curves for mixture F.

CHAPTER 4. CONCLUSIONS AND RECOMMENDED CHANGES TO THE STANDARDS FOR AMPT CYCLIC FATIGUE TESTING

The following summarizes the primary findings and recommended changes to the standards for AMPT cyclic fatigue testing developed as a result of this ruggedness study:

- The statistical significance of the experimental factors evaluated varied across test metrics, mixtures, specimen geometries, and laboratories. In general, air-void content, strain level, and temperature were the most significant experimental factors.

- The effects of statistically significant experimental factors on the material properties determined from the cyclic fatigue tests were found to be very small in some instances. Given the challenges in interpreting the practical significance of the experimental factors on material-level parameters and ultimate pavement performance, the research team used a regression analysis of the FlexPAVE results for a pavement-performance simulation to guide the establishment of tolerance thresholds for the experimental factors. The team used regression model coefficients to estimate the range in values for each factor that will keep the error in predicted cracking below a tolerance threshold of ±10-percent error when all experimental factors are at an ideal experimental condition. Except for air-void content and loading-platen angle, the calculated tolerance ranges were all wider than the experimental levels in the ruggedness evaluation. Consequently, it is recommended that requirements for rest-period duration, loading-platen diameter, and temperature-equilibration times be relaxed in the standards for AMPT cyclic fatigue testing. While specimen height was statistically insignificant, a larger tolerance range for specimen height is not possible given the actuator travel distance and platen thickness in at least one AMPT currently manufactured and used in the United States.

- The tolerance analysis conducted using a regression analysis of the FlexPAVE simulation results suggests that the tolerance range for air-void content should retain the requirement of ±0.5 percent that is included in AASHTO TP 107-18.[1] The same requirement should be added to AASHTO TP 133-19, which does not specify an air-void tolerance range.[2]

- The results of this study showed that a lack of loading-platen parallelism increased the occurrence of specimen end failures in cyclic fatigue tests. The use of shims, placed between the loading platen and AMPT platen prior to tightening the attachment bolts, was found to be an effective strategy to mitigate strain development due to tightening prior to testing when the platens are not parallel. Consequently, end failures are generally avoided when shims are used. The use of a ball bearing was less effective at mitigating strain development upon tightening attachment bolts and is, therefore, not recommended for use in practice.

- A heat transfer model was successfully calibrated and used to establish temperature-conditioning times for the AMPT cyclic fatigue tests; this analysis was based on requiring the specimen to reach ±0.5°C of the target temperature at its center.

- The research team conducted a linear regression analysis of the ruggedness results to relate *DMR* to air-void content and temperature. Acceptance thresholds for air void and temperature were used within the regression equation to establish the acceptable *DMR* limits. The results suggest that the acceptable *DMR* range is 0.85 to 1.15.

- The FlexPAVE analysis results demonstrate that the results of cyclic fatigue testing are sensitive to the number of cycles to failure range. However, a robust means to refine the allowable number of cycles to failure range using the ruggedness test results was not identified. Consequently, variability in the test results associated with the allowable number of cycles to failure range will be embedded into the precision and bias statements developed in the ILS.

- The research team proposes the following changes to the standards for AMPT cyclic fatigue testing based on the previously listed findings:

  o Impose a minimum rest-period duration of 5 min between the dynamic modulus fingerprint loading and cyclic fatigue testing for both large and small specimens. The existing standard specifies a minimum of 20 min of rest.

  o Permit an allowable range in loading platen diameter from 100 mm to 105 mm for large specimens and 37.8 mm to 39.2 mm for small specimens. AASHTO TP 107-18 and AASHTO TP 133-19 specify that loading platens shall be 100 to 105 mm and 37 to 39 mm, respectively.[1,2]

  o Place a maximum limit on the allowable gap between the loading platen and AMPT platen of 1 mm for large specimens and 0.8 mm for small specimens to address loading-platen parallelism. For smaller gaps, the use of shims is required to mitigate strain development induced by tightening the attachment bolts. The existing standards specify the use of a ball bearing in an attempt to account for loading eccentricity with no maximum limit on the gap between the loading platen and AMPT platen.

  o Mandate the temperature conditioning times summarized in table 20 and table 21 for large and small specimens, respectively. The existing standards specify longer conditioning times.

  o Implement an allowable range in *DMR* values from 0.85 to 1.15 for both large and small specimens. The existing standards specify an allowable range in *DMR* values from 0.9 to 1.1.

  o Place a maximum limit of 10 percent on the average actuator standard error in cycles two through five of the cyclic fatigue testing to ensure proper PID settings. The existing standards do not include specifications to ensure proper PID settings.

APPENDIX

PROFICIENCY EXPERIMENTS

Table 26 shows the dynamic modulus testing conditions for the proficiency experiments. As outlined in the main body of the report, dynamic modulus testing on large specimens was conducted in accordance with AASHTO T 378-17, and dynamic modulus testing on small specimens was conducted in accordance with AASHTO TP 132-19.[6,11] Test results obtained at 45°C and 0.01 Hz were neglected from analyses because the measured dynamic modulus values were below the calibration limit of the AMPT. Each laboratory conducted three dynamic modulus tests for each specimen geometry. All cyclic fatigue tests were conducted at 21°C, selected in accordance with AASHTO TP 107-18 and TP 133-19.[1,2] Labs A and B conducted four cyclic fatigue tests using each specimen geometry. The two laboratories used the same input strain levels. In addition, lab A conducted two additional cyclic fatigue tests for each test geometry to select the strain levels required for the ruggedness testing.

Table 26. Proficiency dynamic modulus testing conditions.

Factor	Large Specimen	Small Specimen
Temperatures	4°C, 20°C, and 45°C	4°C, 20°C, and 40°C
Frequencies	10 Hz, 1 Hz, and 0.1 Hz (+0.01 Hz at 45°C)	25 Hz, 10 Hz, 5 Hz, 1 Hz, 0.5 Hz, and 0.1 Hz
Strain level	75 to 125 microstrains	50 to 75 microstrains

The results of the dynamic modulus pilot tests for large and small specimens are shown in figure 89 through figure 92. Within-laboratory and between-laboratory variability is low for both large- and small-specimen geometries except for the phase-angle results at the low reduced frequencies, corresponding to the highest test temperature.
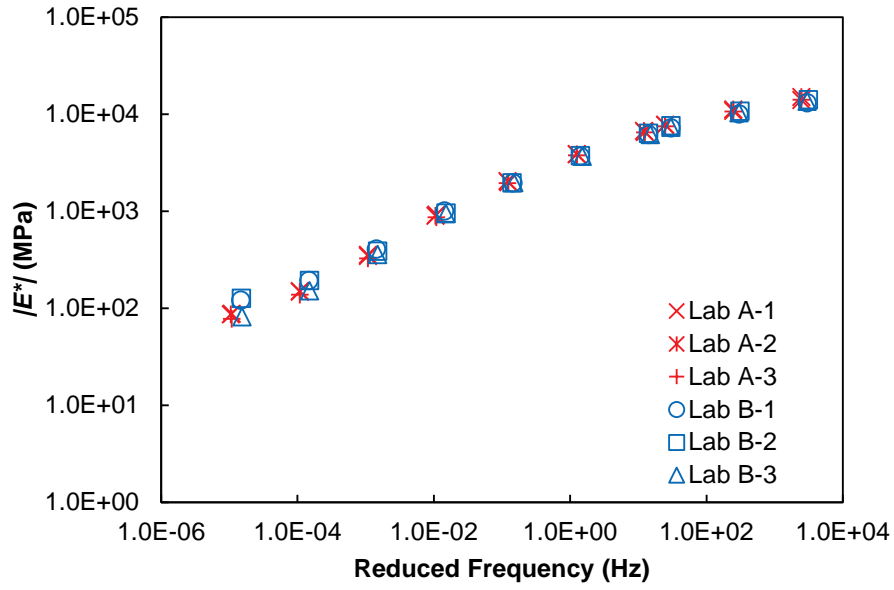
Source: FHWA.

Figure 89. Graph. Dynamic modulus mastercurve for pilot testing a large specimen.
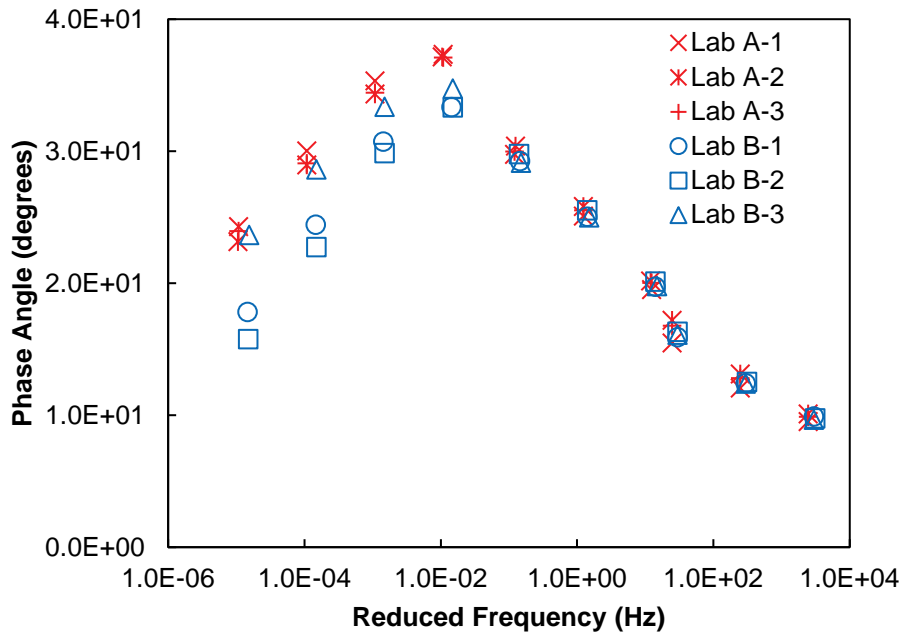
Source: FHWA.

Figure 90. Graph. Phase-angle mastercurve for pilot testing a large specimen.
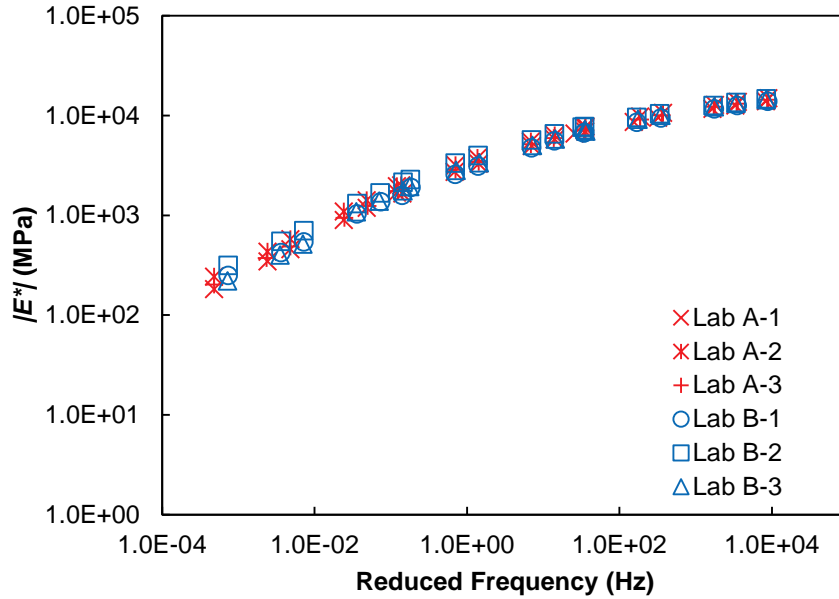
Figure 91. Graph. Dynamic modulus mastercurve for pilot testing a small specimen.
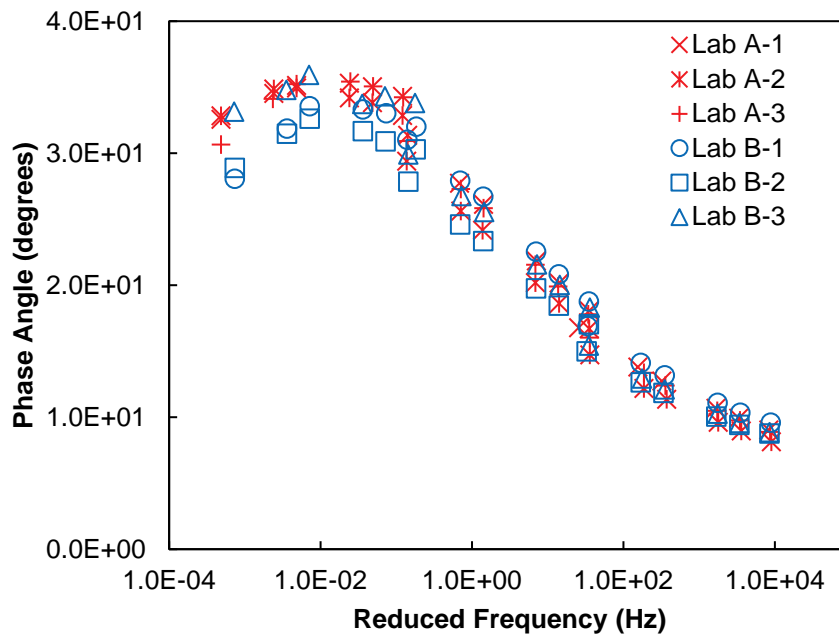
Figure 92. Graph. Phase-angle mastercurve for pilot testing a small specimen.

To further evaluate the reproducibility of the pilot test results, the research team compared the differences between the average dynamic modulus and phase-angle values obtained from the two laboratories against the acceptable thresholds included in the reproducibility statement within AASHTO T 378-17. The team also used an analysis of variance to determine the statistical significance of observed differences between the two laboratories. To evaluate whether the

variance in the results between the two labs was statistically equal, the team conducted F-tests, which determined that the variance was equal with 98-percent confidence in all but one test condition (i.e., 20°C, 0.1 Hz for large specimens). Therefore, an equal variance Student's t-test was conducted to evaluate the significance of differences in the average test results between the two labs.

The large-specimen analysis results for dynamic modulus and phase-angle values are shown in table 27 and table 28, respectively. The small-specimen analysis results for dynamic modulus and phase-angle values are shown in table 29 and table 30, respectively. Test conditions where the differences between the average results of the two labs exceed the threshold specified in AASHTO T 378-17 are indicated with superscripts and footnotes in the tables. At test temperatures of 4°C and 20°C, statistical analysis results indicate that average results are statistically equivalent and AASHTO T 378-17 requirements are met. Table 28 and table 30 demonstrate that the differences in the average phase results of the two laboratories exceed the thresholds specified in AASHTO T 378-17 at the highest test temperature in many instances; however, in all but one instance (i.e., 45°C, 10 Hz for large specimens), the t-test results indicate that the average results of the two laboratories are statistically equal. Table 29 shows that the differences in the average dynamic modulus results of the two laboratories exceed the thresholds specified in AASHTO T 378-17 at three frequencies despite the t-test results indicating that the average results are statistically equal. The reproducibility statement included in AASHTO T 378-17 was developed based on the results of large specimen interlaboratory testing and may not translate directly to small-specimen testing, especially at high test temperatures where the loads used in testing may be near or below the calibrated limit of the AMPT.

The purpose of pilot testing was to determine the agreement of test results between labs A and B for the ruggedness evaluation of the AMPT cyclic fatigue tests. Therefore, the ultimate judgment of the reproducibility in the dynamic modulus pilot test results was based on the combined effects of dynamic modulus and variability in AMPT cyclic fatigue testing on the damage-characteristic curve and failure-criterion results.

Table 27. Reproducibility of dynamic-modulus results for a large specimen.

Temp. (°C)	Freq. (Hz)	Lab B |E*| (MPa)	Lab A |E*| (MPa)	Average |E*| (MPa)	Difference Between Lab A and B (MPa)	Difference of Average (%)	AASHTO T 378 Acceptable Difference (%)	p Value (F-test)	p Value (t-test)
4	10.00	13,581	14,384	13,983	803	5.7	6	0.777	0.129
4	1.00	10,383	10,884	10,634	501	4.7	6	0.960	0.113
4	0.10	7,391	7,680	7,536	290	3.8	8	0.813	0.105
20	10.00	6,235	6,572	6,404	337	5.3	8	0.915	0.023
20	1.00	3,699	3,836	3,768	136	3.6	10	0.623	0.024
20	0.10	1,967	1,994	1,980	27	1.4	13	0.010	0.401
45	10.00	970	895	932	75	8.1	16	0.898	0.043
45	1.00	385	342	363	43	11.7	21	0.534	0.061
45	0.10	180	146	163	34	20.8	27	0.133	0.076
45	0.01	110	84	97	26	26.7	N/A	0.100	0.151

Freq. = frequency; Temp. = temperature.

Table 28. Reproducibility of phase-angle results for a large specimen.

Temp. (°C)	Freq. (Hz)	Lab B Phase Angle (Degree)	Lab A Phase Angle (Degree)	Average Phase Angle (Degree)	Difference Between Lab A and B (Degree)	Difference of Average (Degree)	AASHTO T 378 Acceptable Difference (Degree)	p Value (F-test)	p Value (t-test)
4	10.00	9.8	9.8	9.8	0.1	0.1	0.7	0.318	0.771
4	1.00	12.5	12.7	12.6	0.2	0.2	0.7	0.024	0.552
4	0.10	16.1	16.5	16.3	0.4	0.4	0.8	0.089	0.519
20	10.00	19.9	19.9	19.9	0.0	0.0	0.8	0.542	0.913
20	1.00	25.2	25.5	25.3	0.3	0.3	1.1	0.762	0.359
20	0.10	29.4	30.0	29.7	0.6	0.6	1.4	0.910	0.076
45	10.00	33.8	37.2	35.5	3.4	3.4[a]	1.8	0.043	0.002
45	1.00	31.3	34.7	33.0	3.4	3.4[a]	2.4	0.164	0.037
45	0.10	25.3	29.4	27.3	4.1	4.1[a]	3.2	0.070	0.083
45	0.01	19.1	23.8	21.4	4.7	4.7	N/A	0.040	0.120

[a]The difference exceeds the threshold.

Freq. = frequency; Temp. = temperature.

Table 29. Reproducibility of dynamic-modulus results for a small specimen.

Temp. (°C)	Freq. (Hz)	Lab B $|E^*|$ (MPa)	Lab A $|E^*|$ (MPa)	Average $|E^*|$ (MPa)	Difference Between Lab A and B (MPa)	Difference of Average (%)	AASHTO T 378 Acceptable Difference (%)	p Value (F-test)	p Value (t-test)
4	25	14,345	14,413	14,379	68	0.5	6	0.852	0.869
4	10	13,127	13,185	13,156	58	0.4	6	0.885	0.896
4	5	12,183	12,239	12,211	56	0.5	6	0.901	0.901
4	1	10,010	10,078	10,044	69	0.7	6	0.911	0.879
4	0.5	9,114	9,180	9,147	67	0.7	8	0.888	0.886
4	0.1	7,190	7,180	7,185	10	0.1	8	0.814	0.983
20	25	7,166	7,192	7,179	26	0.4	8	0.464	0.942
20	10	5,990	6,016	6,003	26	0.4	8	0.449	0.942
20	5	5,158	5,205	5,182	46	0.9	8	0.454	0.894
20	1	3,486	3,501	3,494	16	0.4	10	0.451	0.960
20	0.5	2,919	2,923	2,921	4	0.1	10	0.507	0.989
20	0.1	1,830	1,804	1,817	26	1.4	13	0.575	0.899
40	25	2,047	1,802	1,925	245	12.7	13	0.637	0.170
40	10	1,480	1,284	1,382	196	14.2[a]	13	0.663	0.186
40	5	1,145	976	1,060	169	15.9[a]	13	0.638	0.184
40	1	587	507	547	80	14.7	16	0.529	0.307
40	0.5	458	384	421	73	17.4	21	0.478	0.235
40	0.1	262	209	235	52	22.3[a]	21	0.575	0.185

[a]The difference exceeds the threshold.

Freq. = frequency; Temp. = temperature.

Table 30. Reproducibility of phase-angle results for a small specimen.

| Temp. (°C)	Freq. (Hz)	Lab B Phase Angle (Degree)	Lab A Phase Angle (Degree)	Average Phase Angle (Degree)	Difference Between Lab A and B (Degree)	Difference of Average (Degree)	AASHTO T 378 Acceptable Difference (Degree)	p Value (F-test)	p Value (t-test)
4	25	9	9	9	0	0.4	0.7	0.950	0.366
4	10	10	10	10	0	0.2	0.7	0.951	0.621
4	5	10	10	10	0	0.2	0.7	0.979	0.662
4	1	12	12	12	0	0.2	0.7	0.978	0.752
4	0.5	13	13	13	0	0.1	0.8	0.946	0.884
4	0.1	16	16	16	0	0.1	0.8	0.955	0.928
20	25	18	18	18	1	0.5	0.8	0.793	0.473
20	10	20	20	20	0	0.2	0.8	0.636	0.822
20	5	21	21	21	0	0.1	0.8	0.558	0.922
20	1	25	25	25	0	0.2	1.1	0.553	0.889
20	0.5	26	27	27	0	0.4	1.1	0.610	0.718
20	0.1	30	31	30	1	1.0	1.4	0.584	0.431
40	25	32	34	33	2	1.8[a]	1.4	0.364	0.190
40	10	33	35	34	2	1.9[a]	1.4	0.287	0.151
40	5	33	35	34	2	2.1[a]	1.4	0.556	0.049
40	1	34	35	35	1	1.1	1.8	0.015	0.392
40	0.5	33	35	34	2	1.8	2.4	0.079	0.167
40	0.1	30	32	31	2	2.0	2.4	0.322	0.312

[a]The difference exceeds the threshold.

Freq. = frequency; Temp. = temperature.

The pilot-test results for AMPT cyclic fatigue testing are shown in figure 93 and figure 94. The dynamic modulus test results from each laboratory were combined with their respective results from AMPT cyclic fatigue testing for analysis. Thus, the results of the AMPT cyclic fatigue testing presented include the combined effects of variability in dynamic modulus and cyclic fatigue testing on the outcomes of the AMPT cyclic fatigue tests (i.e., the damage-characteristic curve and D^R failure criteria).

Figure 93 shows that the damage-characteristic curves obtained from all tests align, indicating good agreement. The D^R criterion is the average reduction in C up to failure. The variability in the D^R results of test replicates can be observed in the plot of the summation of $(1 - C)$ up to the point of failure versus N_f. The D^R data shown in figure 94 include all results for the pilot cyclic fatigue tests; however, one of the tests conducted at lab A on a small specimen yielded a very high N_f because the onspecimen strain in the test was lower than anticipated. Therefore, the D^R results are shown in figure 95 with removal of this long test to allow for easier interpretation of the test variability. The D^R values demonstrate low variability as the data points align to form a line that passes through the origin.

The results demonstrate acceptable variability within and between laboratories in terms of both damage-characteristic curve and failure criterion results based on the lab A research team's past experience with AMPT cyclic fatigue testing. The pilot testing results, therefore, were deemed acceptable, indicating that the laboratories are prepared to begin ruggedness testing.
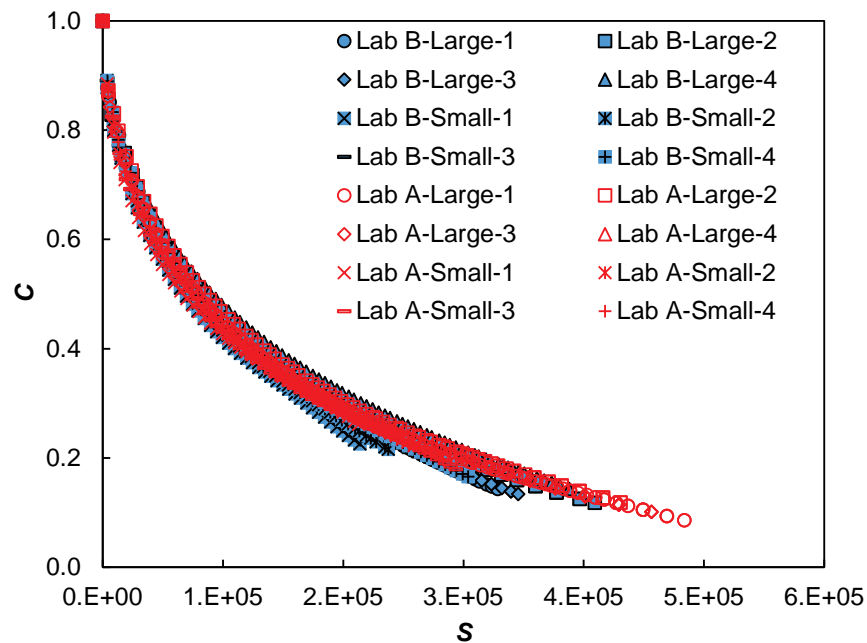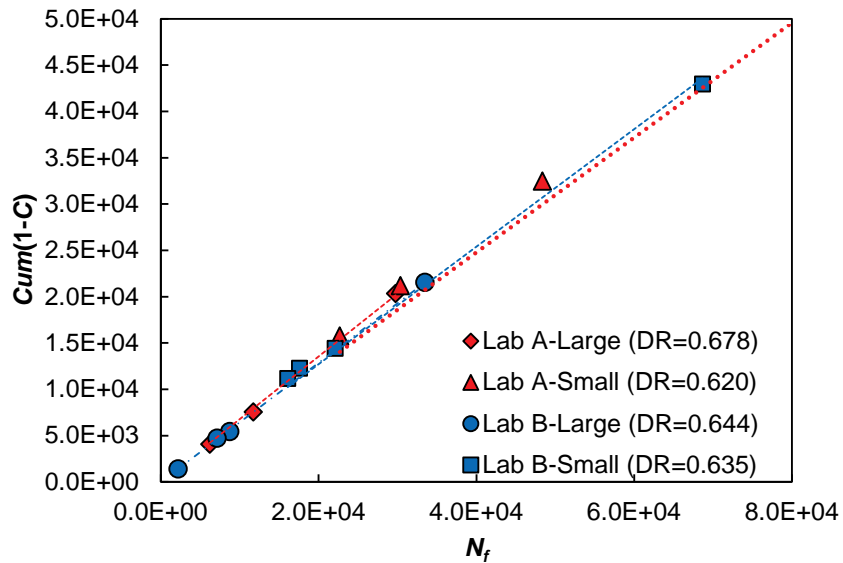

Source: FHWA.

Figure 93. Graph. Damage-characteristic curves for pilot testing.
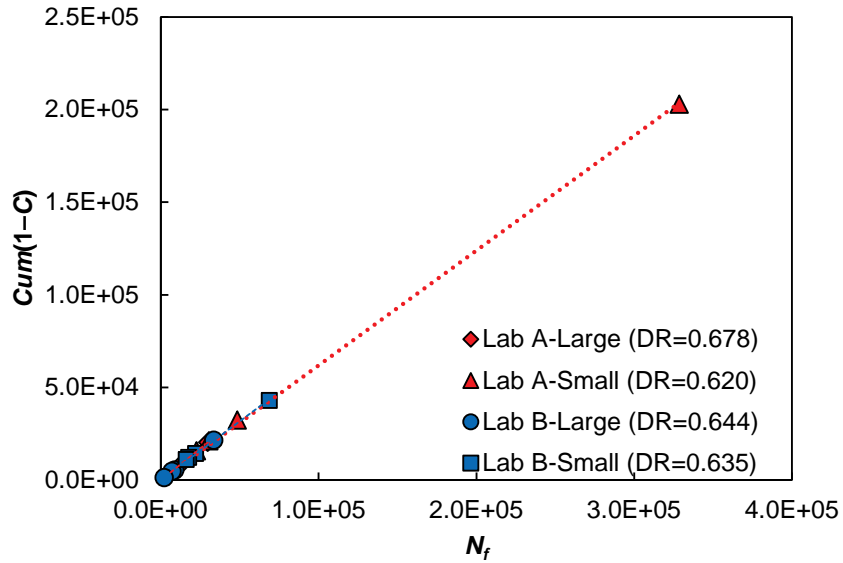
Source: FHWA.

**Figure 94. Graph. D^R failure criterion (all data).**

Source: FHWA.

Figure 95. Graph. D^R failure criterion (excluding long test).

## PILOT-TEST RESULTS FOR STRAIN-LEVEL SELECTION

Strain level selection for the ruggedness evaluation was performed for the 9.5-mm NMAS mixture. The lab A research team used the pilot-test results for the 9.5-mm NMAS mixture from both labs A and B to select the strain levels required for ruggedness testing. It was first found that machine compliance differences led to different relationships between the input strain level and actual onspecimen strain levels for large and small specimens. Furthermore, the machine compliance of lab A's AMPT and lab B's AMPT differ. As a result, relationships between input

strain levels and the measured number of cycles were established for both large and small geometries in each laboratory, as shown in figure 96. Based on the relationships, corresponding strain levels, which target 5,000; 10,000; 20,000; 40,000; and 80,000 cycles to failure, were selected. Note that the actual fatigue lives measured in each ruggedness determination are expected to vary; these strain levels are simply selected to yield general ranges of fatigue lives for the low and high levels. Level 1 strain levels were selected to reflect typical practice, and level 2 strain levels were selected to represent the maximum length of tests that would be tolerated in a lab given practical time constraints. Figure 97 shows the relationships for 19- and 25-mm NMAS mixtures.
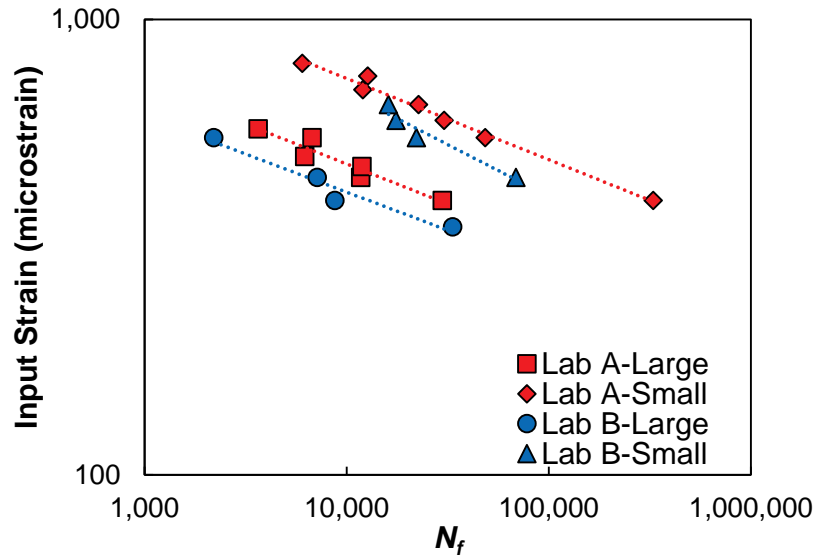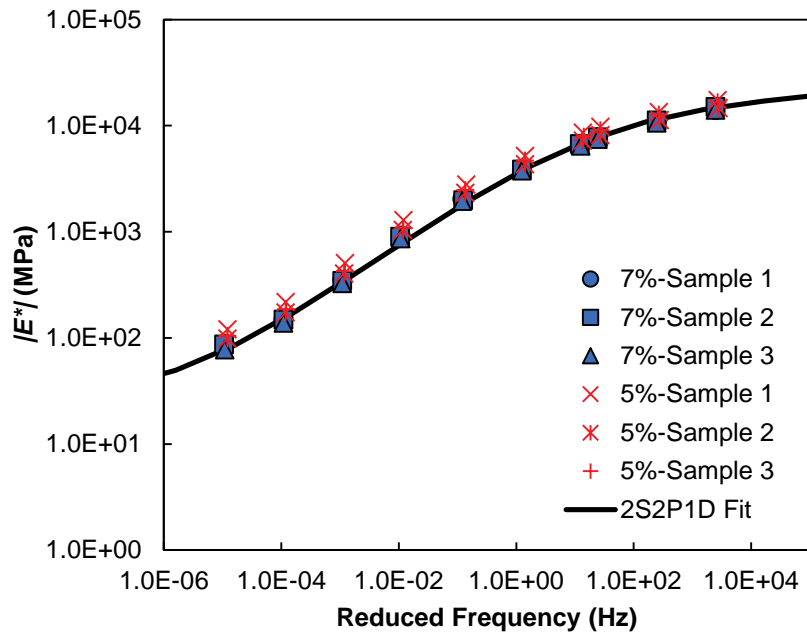

Source: FHWA.

Figure 96. Graph. Determination of strain levels using input strain versus fatigue-life relationship for the 9.5-mm NMAS mixture.
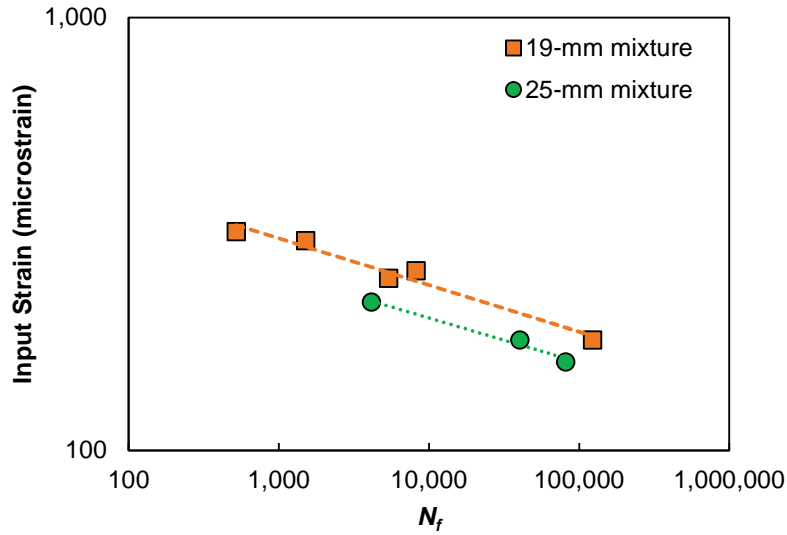
Source: FHWA.

Figure 97. Graph. Determination of strain levels using input strain versus fatigue-life relationship for the 19-mm and 25-mm NMAS mixtures.

DYNAMIC MODULUS TEST RESULTS

Source: FHWA.

Figure 98. Graph. Lab A results of the dynamic modulus test for a large specimen of 9.5-mm NMAS mixture (log–log plot).
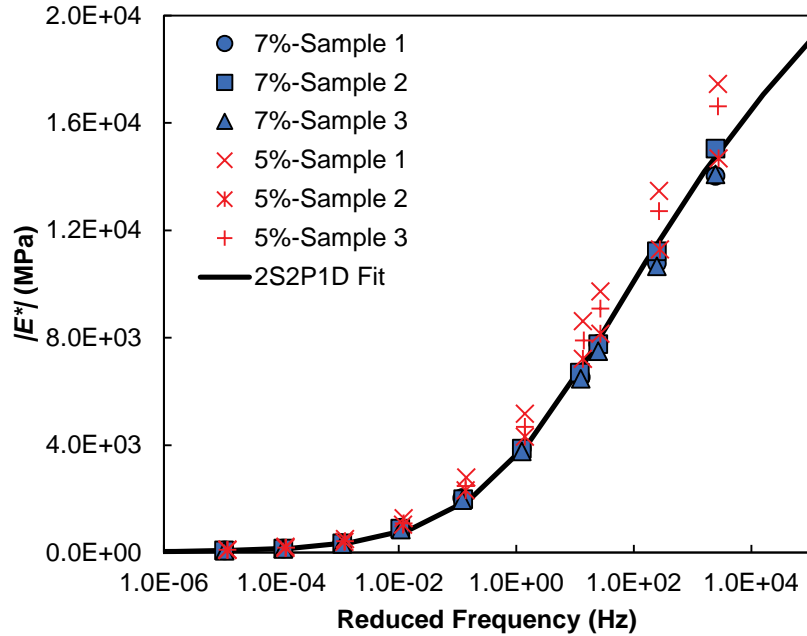
Figure 99. Graph. Lab A results of the dynamic modulus test for a large specimen of 9.5-mm NMAS mixture (semilog plot).

Figure 100. Graph. Lab A results of the dynamic modulus test for a small specimen of 9.5-mm NMAS mixture (log–log plot).

Figure 101. Graph. Lab A results of the dynamic modulus test for a small specimen of 9.5-mm NMAS mixture (semilog plot).
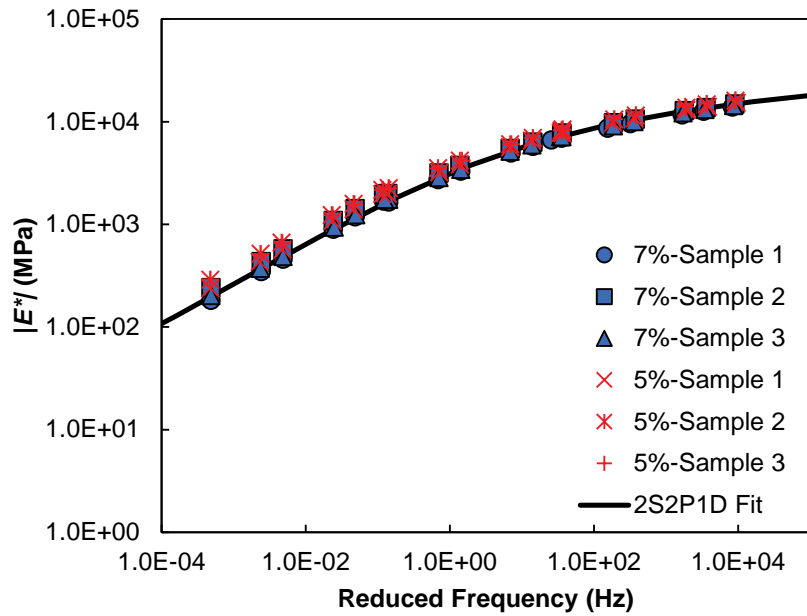
Figure 102. Graph. Lab B results of the dynamic modulus test for a large specimen of 9.5-mm NMAS mixture (log–log plot).

Figure 103. Graph. Lab B results of the dynamic modulus test for a large specimen of 9.5-mm NMAS mixture (semilog plot).
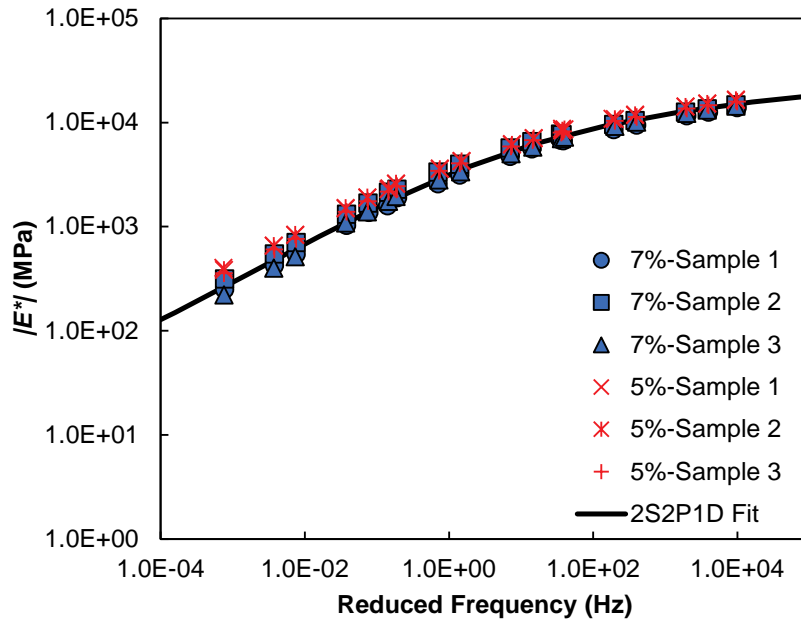
Figure 104. Graph. Lab B results of the dynamic modulus test for a small specimen of 9.5-mm NMAS mixture (log–log plot).
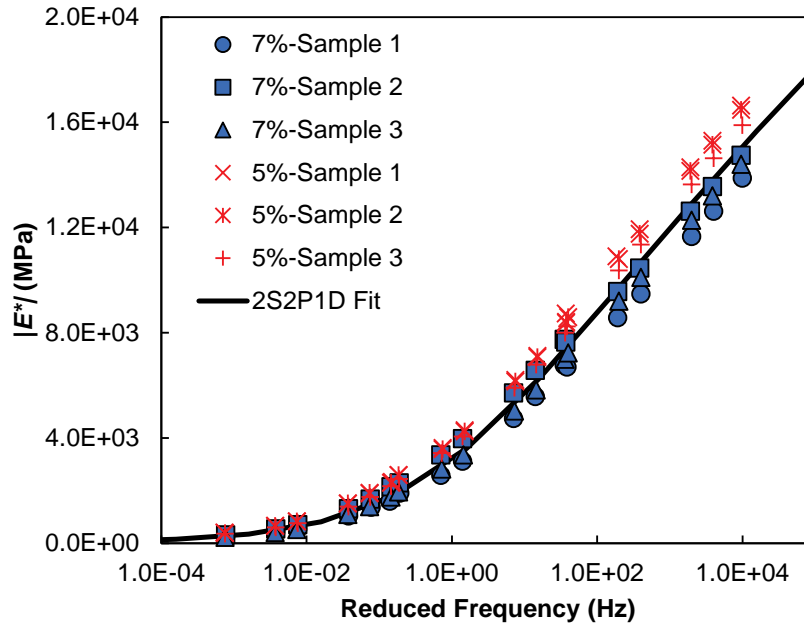
Figure 105. Graph. Lab B results of the dynamic modulus test for a small specimen of 9.5-mm NMAS mixture (semilog plot).
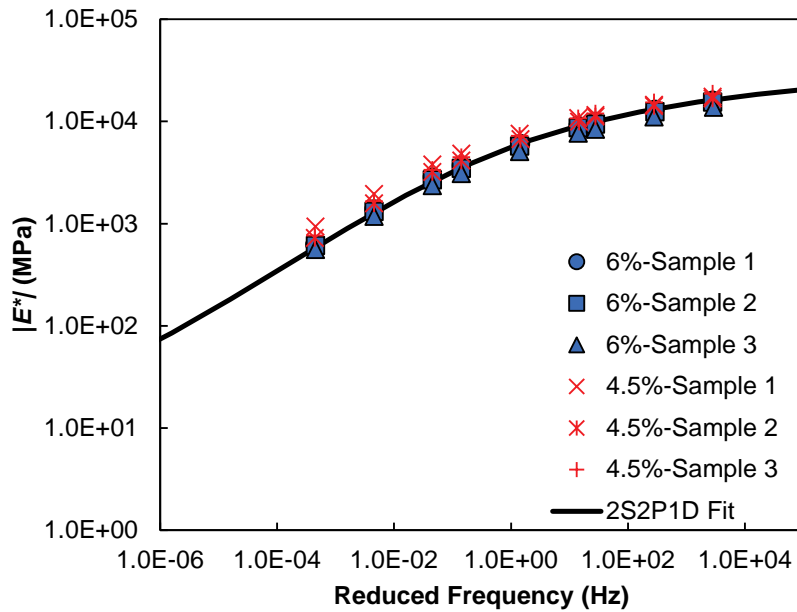
Figure 106. Graph. Results of the dynamic modulus test for a large specimen of 25-mm NMAS mixture (log–log plot).

99
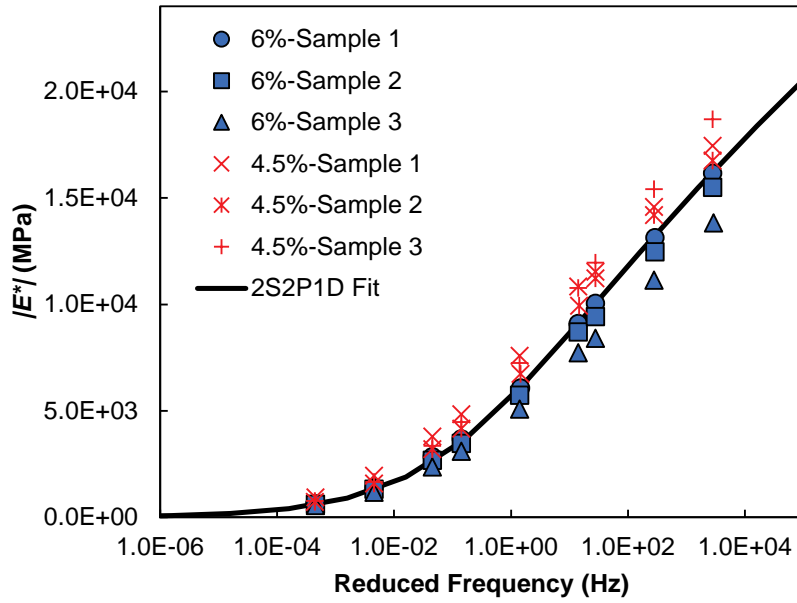
Figure 107. Graph. Results of the dynamic modulus test for a large specimen of 25-mm NMAS mixture (semilog plot).
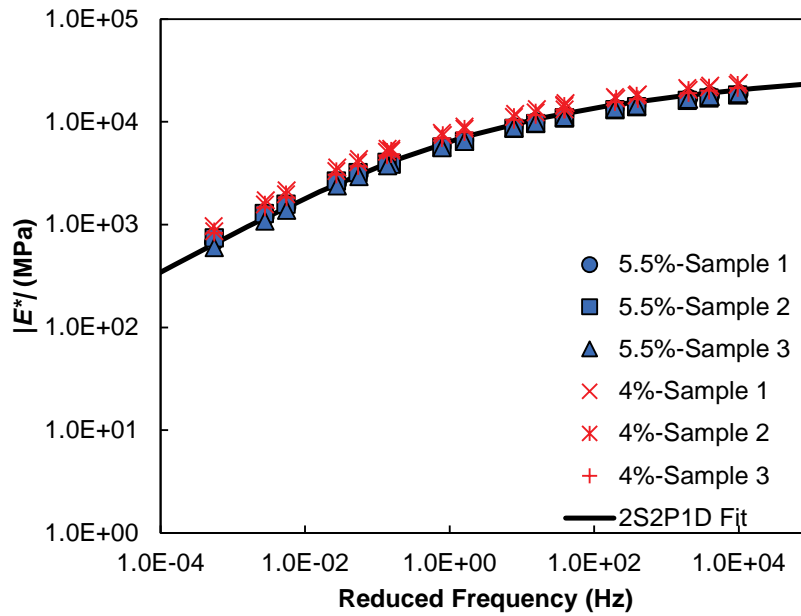
Figure 108. Graph. Results of the dynamic modulus test for a small specimen of 19-mm NMAS mixture (log–log plot).
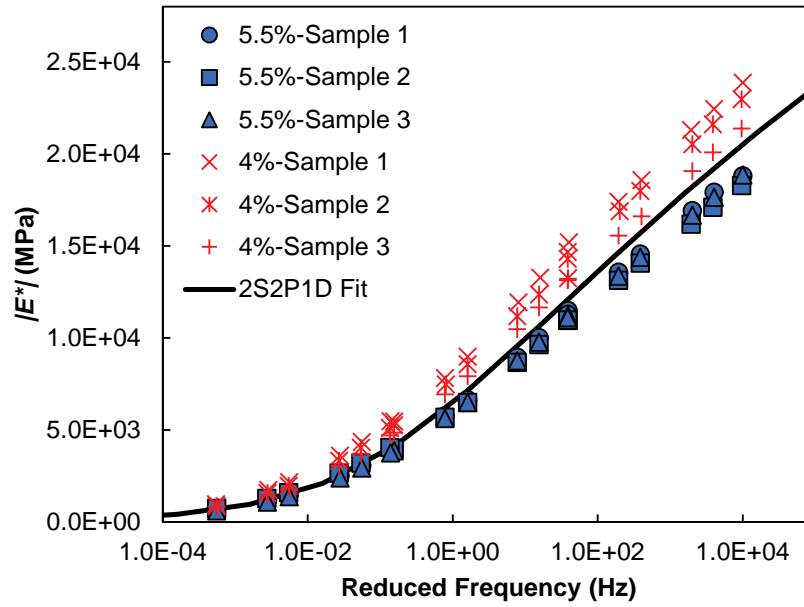
Source: FHWA.

Figure 109. Graph. Results of the dynamic modulus test for a small specimen of 19-mm NMAS mixture (semilog plot).
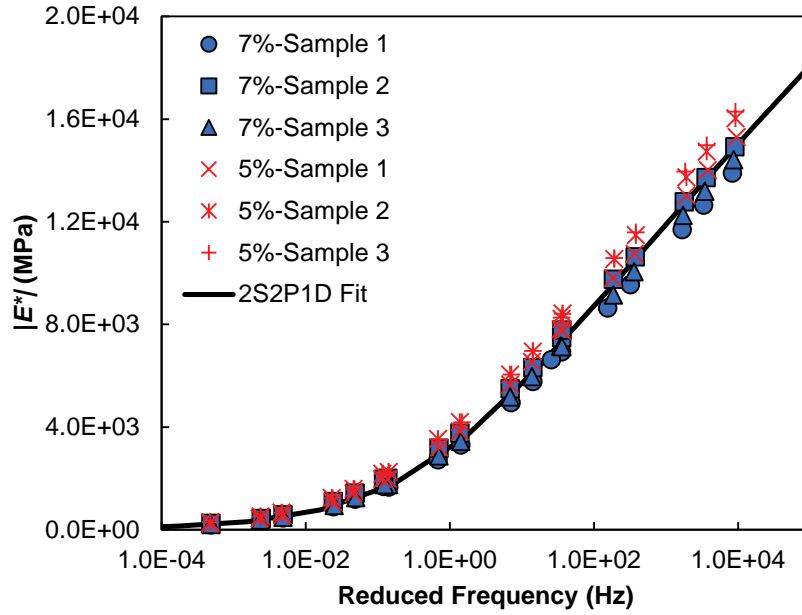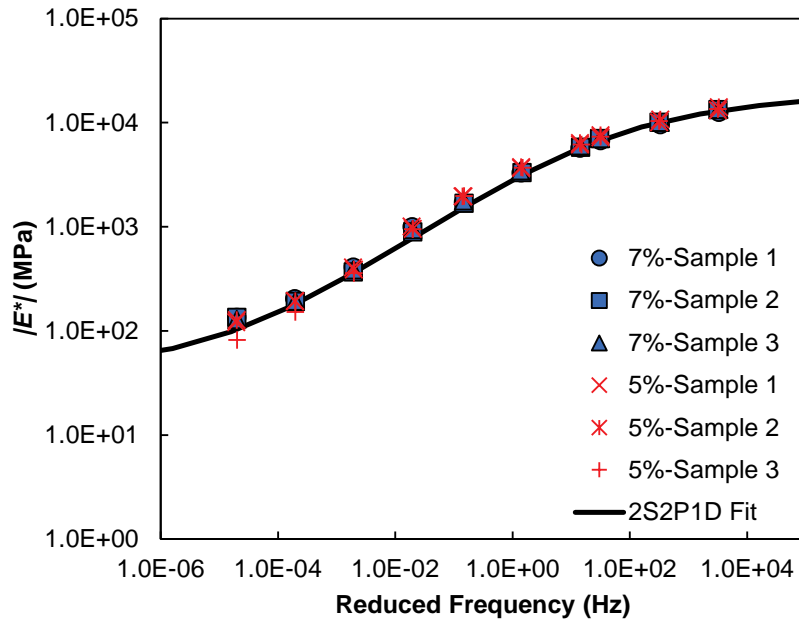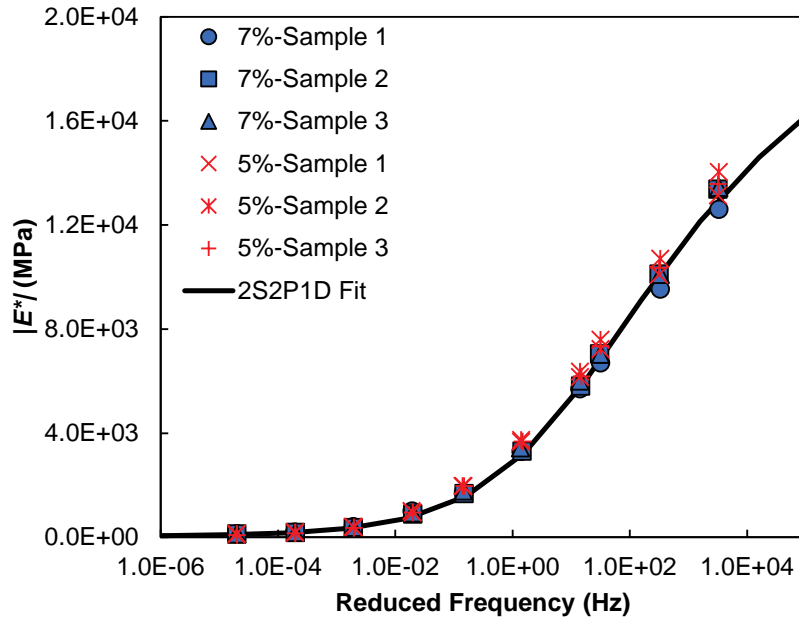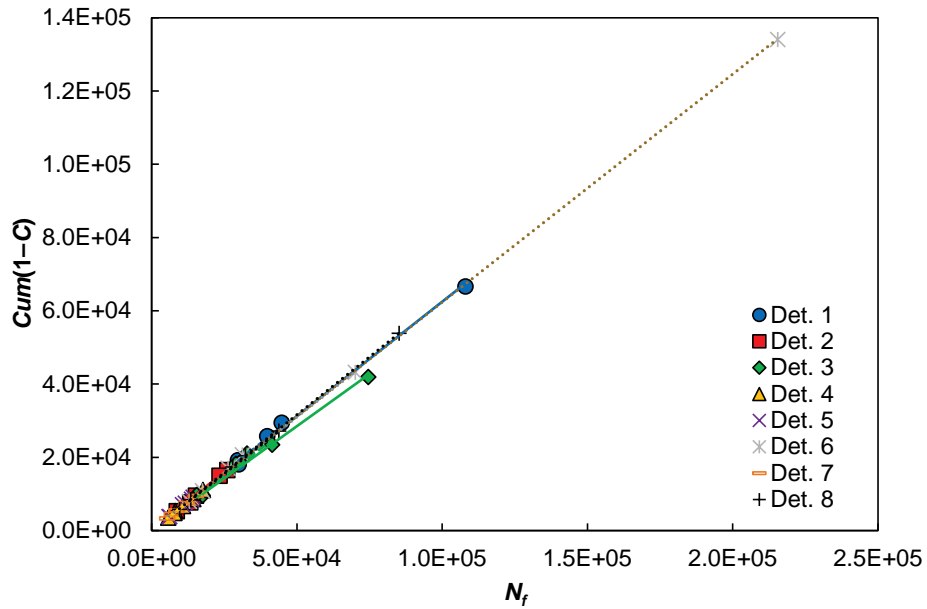
Source: FHWA.

**Figure 110. Graph. D^R failure criterion for lab A resulting from testing a large specimen of 9.5-mm NMAS mixture.**
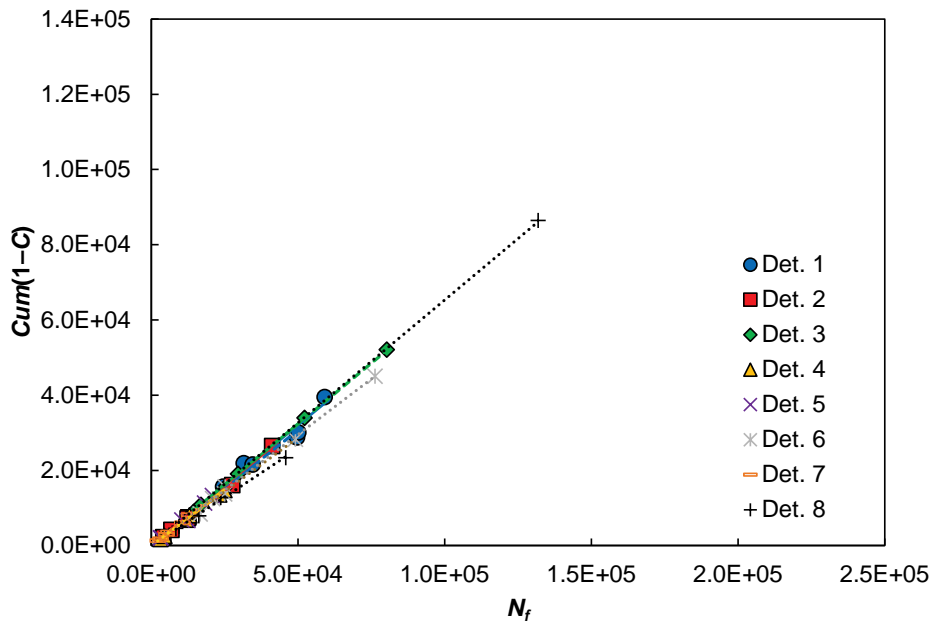
Source: FHWA.

Figure 111. Graph. D^R failure criterion for lab B resulting from testing a large specimen of 9.5-mm NMAS mixture.
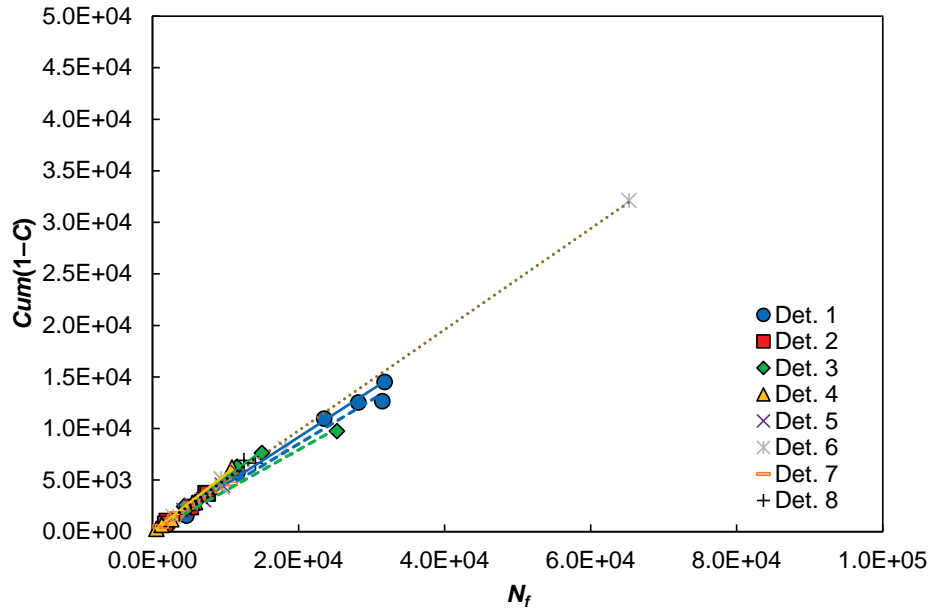
Source: FHWA.

Figure 112. Graph. D^R failure criterion resulting from testing a large specimen of 25-mm NMAS mixture.

Source: FHWA.

Figure 113. Graph. D^R failure criterion for lab A resulting from testing a small specimen of 9.5-mm NMAS mixture.

Source: FHWA.

Figure 114. Graph. D^R failure criterion for lab B resulting from testing a small specimen of 9.5-mm NMAS mixture.
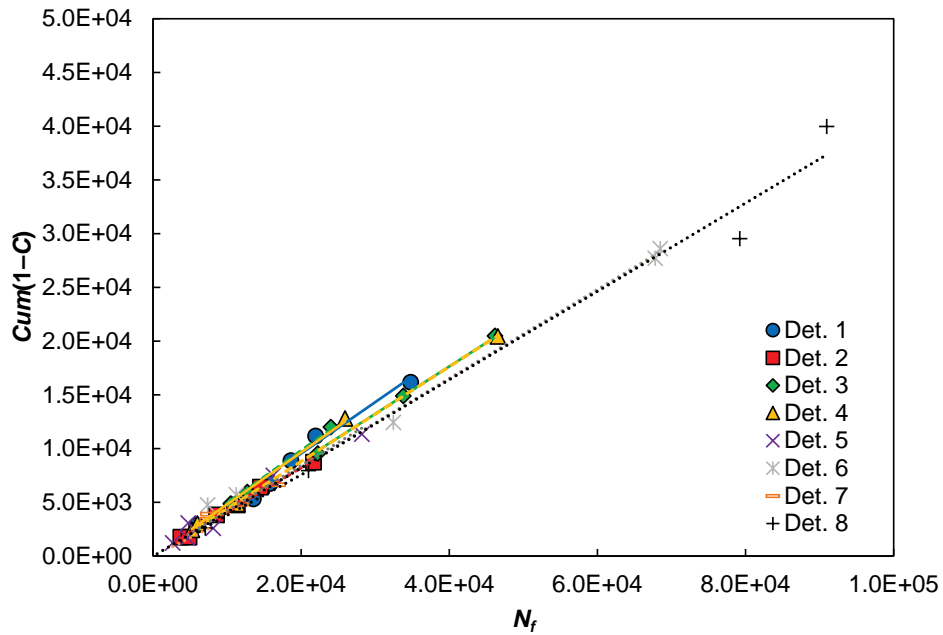
Source: FHWA.

Figure 115. Graph. D^R failure criterion resulting from testing a small specimen of 19-mm NMAS mixture.

STATISTICAL ANALYSIS RESULTS

Table 31. *p* values for the lab A ruggedness evaluation of the 9.5-mm small specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.1869	0.0086	0.0030	0.0012	0.0313	0.0013
Specimen height	0.3690	0.5949	0.9543	0.9244	0.0370	0.0668
Loading-platen parallelism	0.1991	0.1575	0.2785	0.9250	0.8766	0.9160
Difference in diameter of loading platen and specimen	0.4007	0.7850	0.6835	0.9384	0.0160	0.0044
Strain level	0.4285	0.0096	0.0017	0.0009	0.0101	0.0006
Rest period	0.1753	0.2051	0.4688	0.7600	0.2779	0.0638
Temperature	0.4884	0.0531	0.0174	0.0103	0.2327	0.0007

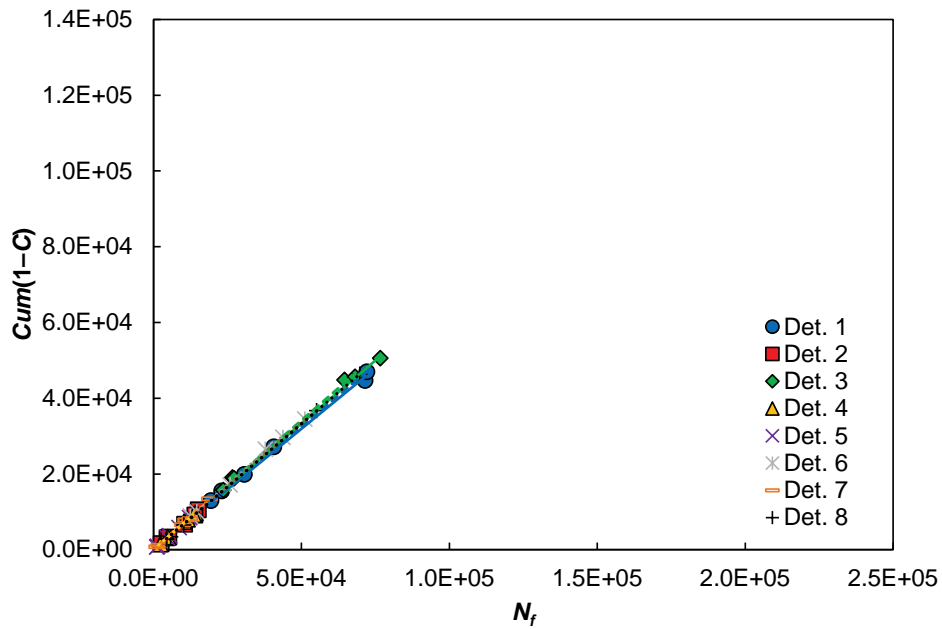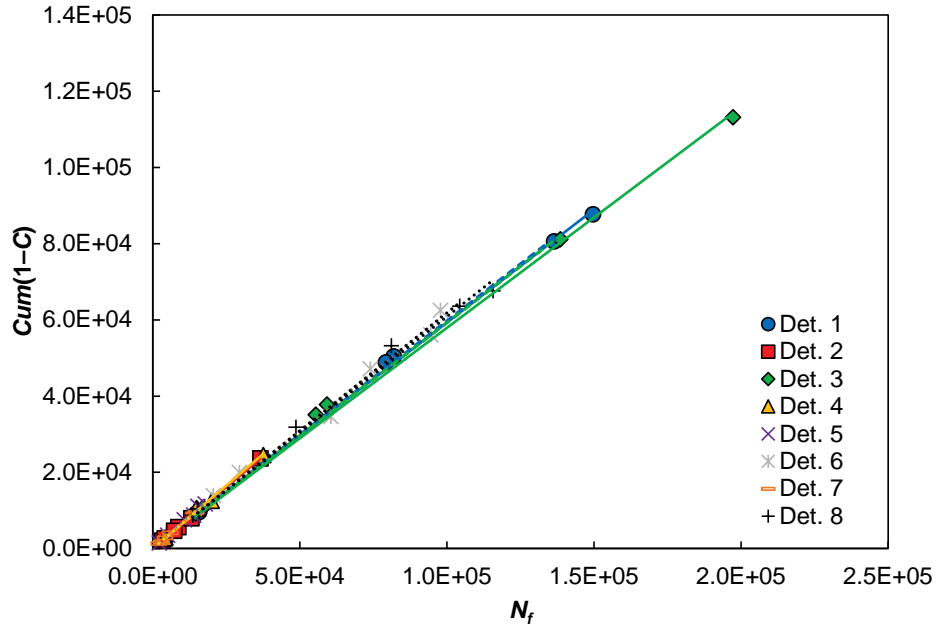**Table 32. *p* values for the lab B ruggedness evaluation of the 9.5-mm small specimen.**

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.3950	0.0110	0.0016	0.0005	0.1399	<0.0000
Specimen height	0.5752	0.9272	0.6691	0.9197	0.0505	0.4210
Loading-platen parallelism	0.2299	0.1468	0.1977	0.9203	0.9738	0.5419
Difference in diameter of loading platen and specimen	0.4194	0.5797	0.9053	0.9346	0.3529	0.0020
Strain level	0.4231	0.0121	0.0017	0.0004	0.3128	0.0000
Rest period	0.1456	0.2198	0.5703	0.7456	0.1583	0.0120
Temperature	0.7123	0.0312	0.0037	0.0063	0.1532	0.1622

Table 33. *p* values for the ruggedness evaluation of the 19-mm small specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.0062	0.1487	0.3119	0.1713	0.0179	0.5282
Specimen height	0.8945	0.4188	0.3462	0.5349	0.2572	0.1595
Loading-platen parallelism	0.1185	0.1909	0.2178	0.2860	0.7069	0.4363
Difference in diameter of loading platen and specimen	0.8805	0.9784	0.9943	0.8084	0.6279	0.6790
Strain level	0.1878	0.3530	0.4114	0.4242	0.6406	0.6116
Rest period	0.6914	0.7893	0.8209	0.8803	0.5607	0.9415
Temperature	0.1363	0.1937	0.2139	0.1358	0.2530	0.0871

Table 34. *p* values for the lab A ruggedness evaluation of the 9.5-mm large specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.1869	0.0086	0.0030	0.0023	0.0163	0.0010
Specimen height	0.3690	0.5949	0.9543	0.9626	0.0065	0.0330
Loading-platen parallelism	0.1991	0.1575	0.2785	0.2107	0.2796	0.1931
Difference in diameter of loading platen and specimen	0.4007	0.7850	0.6835	0.5319	0.0194	0.0789
Strain level	0.4285	0.0096	0.0017	0.0009	0.0143	0.0008
Rest period	0.1753	0.2051	0.4688	0.7812	0.0185	0.1889
Temperature	0.4884	0.0531	0.0174	0.0155	0.2858	0.0408

Table 35. *p* values for the lab B ruggedness evaluation of the 9.5-mm large specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.0001	0.0007	0.0043	0.0035	0.3134	0.9429
Specimen height	0.0108	0.2258	0.8801	0.2655	0.8734	0.9668
Loading-platen parallelism	0.5163	0.0007	0.0001	0.0005	0.1548	0.0297
Difference in diameter of loading platen and specimen	0.0220	0.2492	0.7919	0.7815	0.8382	0.9924
Strain level	0.0025	0.0001	0.0000	0.0000	0.8794	0.2251
Rest period	0.7940	0.4321	0.3792	0.8846	0.8355	0.9942
Temperature	0.0006	0.0866	0.9214	0.6243	0.4835	0.3350

Table 36. *p* values for the ruggedness evaluation of the 25-mm large specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.2390	0.1251	0.1592	0.0001	0.9700	0.0354
Specimen height	0.4263	0.7960	0.9838	0.3710	0.7706	0.9812
Loading-platen parallelism	0.2523	0.4704	0.6902	0.5049	0.1151	0.2064
Difference in diameter of loading platen and specimen	0.4571	0.8951	0.8847	0.5762	0.3096	0.2599
Strain level	0.4838	0.1329	0.1233	0.1071	0.2211	0.1230
Rest period	0.2264	0.5222	0.7943	0.6042	0.8291	0.8159
Temperature	0.5406	0.2996	0.3094	0.0048	0.7556	0.1916

Table 37. Main effects for the lab A ruggedness evaluation of the 9.5-mm small specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.0050	0.0115	0.0164	−3551	0.014	3.793
Specimen height	0.0033	0.0018	−0.0002	646	0.013	1.667
Loading-platen parallelism	−0.0048	−0.0051	−0.0043	−199	0.001	0.086
Difference in diameter of loading platen and specimen	0.0030	0.0009	−0.0016	2221	−0.017	−3.092
Strain level	0.0029	0.0113	0.0182	−3642	0.018	4.284
Rest period	−0.0051	−0.0045	−0.0028	1045	−0.006	−1.691
Temperature	0.0025	0.0074	0.0114	−5706	0.007	4.176

Table 38. Main effects for the lab B ruggedness evaluation of the 9.5-mm small specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.0087	0.0255	0.0391	−8,963	0.022	6.222
Specimen height	0.0004	0.0029	0.0052	−1,513	−0.002	0.407
Loading-platen parallelism	0.0026	−0.0004	−0.0035	111	−0.003	0.306
Difference in diameter of loading platen and specimen	0.0013	−0.0044	−0.0098	1,757	−0.010	−2.163
Strain level	0.0147	0.0242	0.0296	−7,812	0.053	9.349
Rest period	−0.0025	−0.0010	0.0012	−203	0.013	1.552
Temperature	−0.0164	−0.0054	0.0084	−859	0.004	0.739

Table 39. Main effects for the ruggedness evaluation of the 19-mm small specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.0216	0.0353	0.0408	−6,957	0.013	−0.513
Specimen height	0.0008	0.0187	0.0378	−2,045	−0.016	1.207
Loading-platen parallelism	−0.0099	−0.0315	−0.0507	4,595	−0.005	-0.637
Difference in diameter of loading platen and specimen	−0.0009	−0.0006	0.0003	269	0.012	−0.334
Strain level	0.0082	0.0216	0.0327	−3,830	0.013	0.411
Rest period	0.0023	0.0060	0.0088	−915	0.013	0.059
Temperature	0.0094	0.0313	0.0512	−5,701	0.005	1.516

Table 40. Main effects for the lab A ruggedness evaluation of the 9.5-mm large specimen.

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.0050	0.0115	0.0164	−6,193	0.013	2.685
Specimen height	0.0033	0.0018	−0.0002	68	−0.016	−1.372
Loading-platen parallelism	−0.0048	−0.0051	−0.0043	1,923	−0.005	−0.758
Difference in diameter of loading platen and specimen	0.0030	0.0009	−0.0016	−923	0.012	1.074
Strain level	0.0029	0.0113	0.0182	−7,298	0.013	2.803
Rest period	−0.0051	−0.0045	−0.0028	406	0.013	0.766
Temperature	0.0025	0.0074	0.0114	−4,328	0.005	1.299

Table 41. Main effects for the lab B ruggedness evaluation of the 9.5-mm large specimen.

| Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.0133	0.0175	0.0182	−4,779	−0.022	0.138
Specimen height	0.0061	0.0041	0.0007	−1,398	−0.003	−0.080
Loading-platen parallelism	0.0012	0.0175	0.0324	−6,583	0.032	4.933
Difference in diameter of loading platen and specimen	0.0052	0.0038	0.0012	−335	−0.004	-0.018
Strain level	0.0082	0.0249	0.0388	−9,380	−0.003	2.456
Rest period	0.0005	0.0026	0.0041	175	−0.004	−0.014
Temperature	−0.0105	−0.0061	0.0004	595	−0.015	−1.917

**Table 42. Main effects for the ruggedness evaluation of the 25-mm large specimen.**

Factor	C at Low S Value	C at Medium S Value	C at High S Value	Area Above C versus S Curve	D^R	S_{app}
Air voids	0.0329	0.0509	0.0705	−10,918	0.013	1.537
Specimen height	−0.0038	−0.0065	−0.0100	1,529	−0.016	0.015
Loading-platen parallelism	0.0038	0.0042	0.0038	−1,126	−0.005	−0.836
Difference in diameter of loading platen and specimen	0.0027	0.0001	−0.0048	940	0.012	−0.737
Strain level	0.0053	0.0110	0.0186	−2,928	0.013	1.048
Rest period	0.0011	0.0029	0.0054	−871	0.013	0.146
Temperature	0.0099	0.0206	0.0350	−6,234	0.005	0.867

THRESHOLD ANALYSIS MODELS
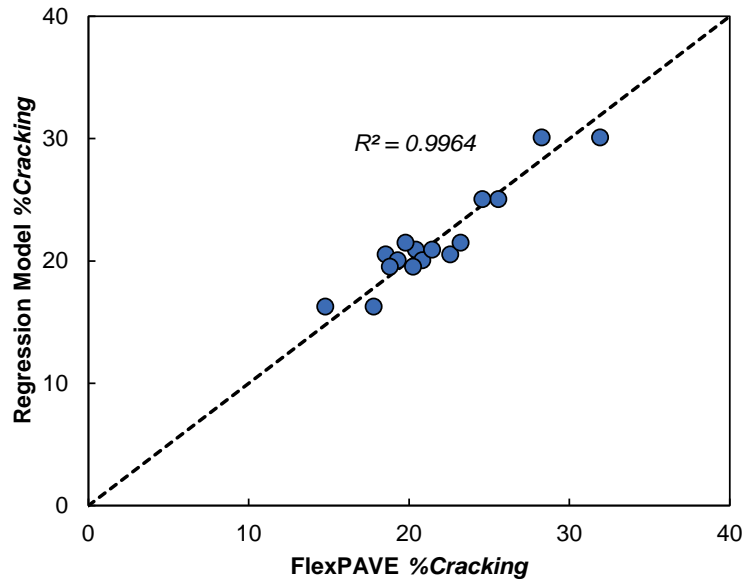
Source: FHWA.

Figure 116. Graph. Regression model versus FlexPAVE *%Cracking* on determination-by-determination basis resulting from lab A testing a large specimen of 9.5-mm NMAS mixture.
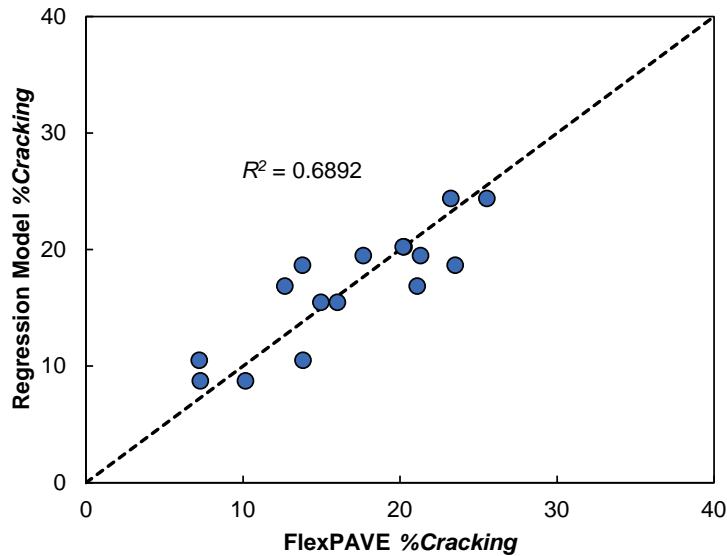
Source: FHWA.

Figure 117. Graph. Regression model versus FlexPAVE *%Cracking* on determination-by-determination basis resulting from lab B testing a large specimen of 9.5-mm NMAS mixture.

$R^2 = 0.6523$

Figure 118. Graph. Regression model versus FlexPAVE *%Cracking* on determination-by-determination basis resulting from testing a large specimen of 25-mm NMAS mixture.

$R^2 = 0.9339$

Figure 119. Graph. Regression model versus FlexPAVE *%Cracking* on determination-by-determination basis resulting from lab A testing a small specimen of 9.5-mm NMAS mixture.
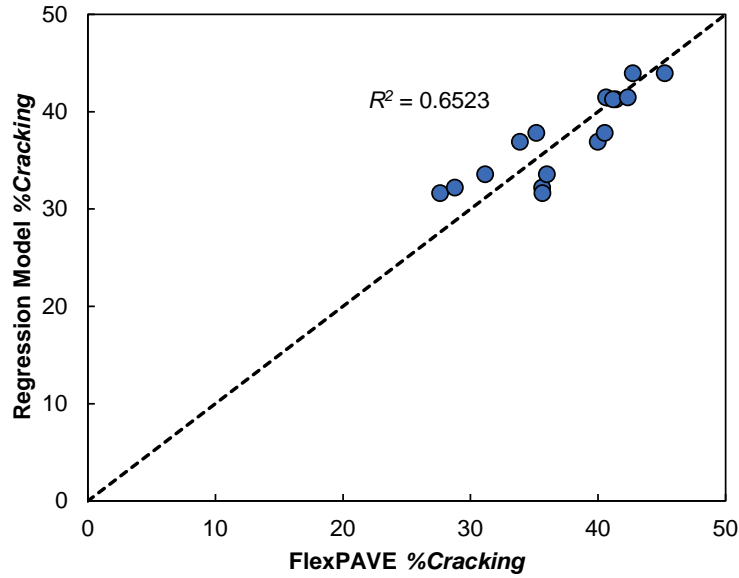
Source: FHWA.

Figure 120. Graph. Regression model versus FlexPAVE *%Cracking* on determination-by-determination basis resulting from lab B testing a small specimen of 9.5-mm NMAS mixture.
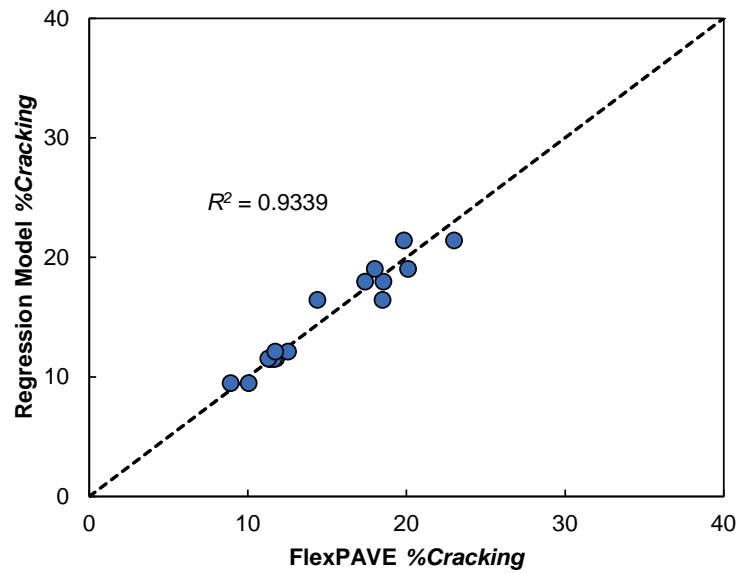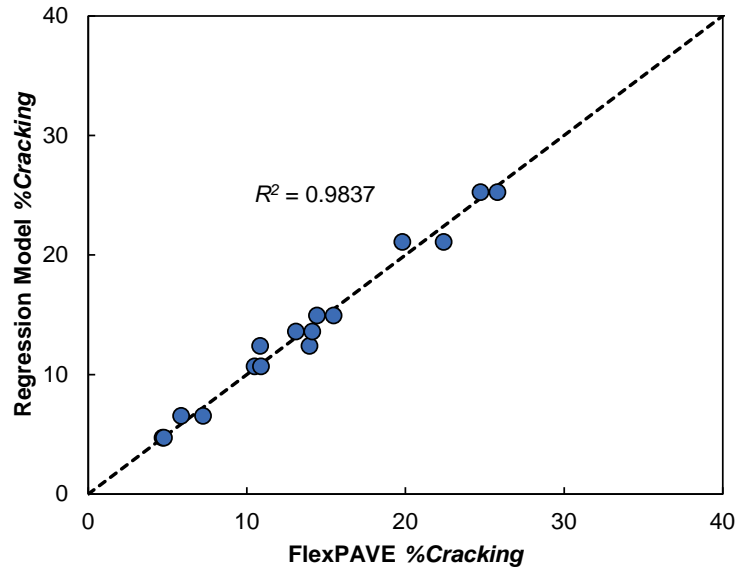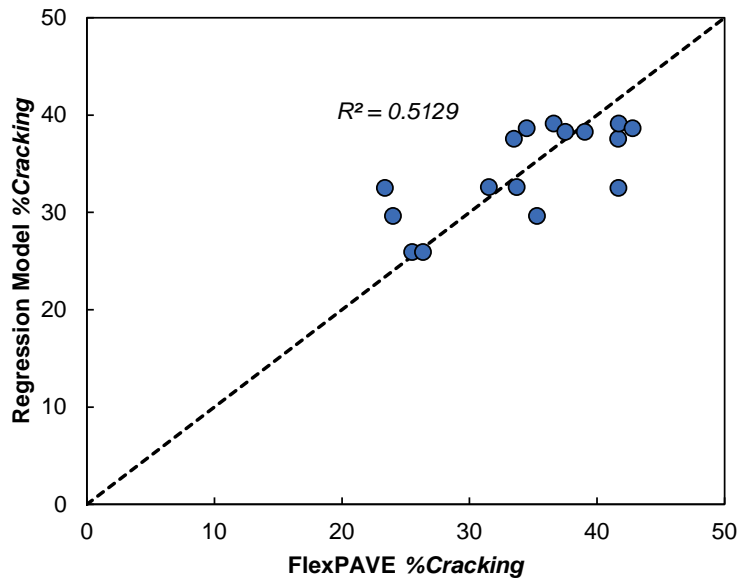

Source: FHWA.

Figure 121. Graph. Regression model versus FlexPAVE *%Cracking* on determination-by-determination basis resulting from testing a small specimen of 19-mm NMAS mixture.

ALLOWABLE GAP CALCULATIONS

This section presents the equations used to calculate the allowable gaps for small- and large-specimen geometries. In figure 55, d, g^e, and g^m represent loading-platen angle, effective gap, and measured gap, respectively. The effective gap is the gap that affects strain development upon

111

tightening the attachment bolts, and the measured gap will be determined by the loading-platen diameters. The effective and measured gap equations are presented in figure 122.

$$g^e = \left\{\varphi_{platen} - \left(\frac{\varphi_{platen} - \varphi_{specimen}}{2}\right)\right\} \times \arctan\left(\frac{\pi}{180°}q\right)$$

$$g^m = \varphi_{platen} \times \arctan\left(\frac{\pi}{180}d\right)$$

Figure 122. Equation. Effective and measured gap equations.

Where:

φ_{platen} = diameter of the loading platen (mm).
$\varphi_{specimen}$ = diameter of the test specimen (mm).
q = angle of the loading platen (degree).

When d is equal to 0.6 degrees, which is the threshold established from the ruggedness evaluation, the effective and measured gaps for a loading platen with an 80-mm diameter will be approximately 0.6 and 0.8 mm, respectively. In the ruggedness study, a loading platen with an 80-mm diameter was used for small-specimen testing, but a loading platen with a 130-mm diameter is considered feasible as this is the diameter of the AMPT platen. For loading platens with a 130-mm diameter, the measured gap should be calculated by using the same effective gap, which is 0.6 mm, from the loading platen with an 80-mm diameter. Figure 123 presents the calculation of measured gap for loading platens with a 130-mm diameter, which is approximately 1.0 mm. Similar to the calculation for the small-specimen geometry, the measured gap for the large-specimen geometry is determined and the threshold is 1 mm. Consequently, the maximum allowable gaps for the large- and small-specimen geometry are 1 mm and 0.8 mm, respectively.

$$g_1^e = \left\{80 - \left(\frac{80-38}{2}\right)\right\} \times \arctan\left(\frac{\pi}{180} \times 0.6\right) = 0.6$$

$$g_1^e = g_2^e \left\{130 - \left(\frac{130-38}{2}\right)\right\} \times \arctan\left(\frac{\pi}{180} \times d_2\right) = 0.6$$

$$\therefore d_2 = 0.42°$$

$$g_2^m = 130 \times \arctan\left(\frac{\pi}{180} \times 0.42\right) = 1.0$$

Figure 123. Equation. Measured gap for loading platens with a 130-mm diameter.

REFERENCES

1. AASHTO TP 107-18. *Determining the Damage Characteristic Curve and Failure Criterion Using the Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test*. AASHTO, 2018.

2. AASHTO TP 133-19. *Determining the Damage Characteristic Curve and Failure Criterion Using Small Specimens in the Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test*. AASHTO, 2019.

3. ASTM E1169-14. *Standard Practice for Conducting Ruggedness Tests*. ASTM International, 2014.

4. ASTM C1067-12. *Standard Practice for Conducting a Ruggedness Evaluation or Screening Program for Test Methods for Construction Materials*. ASTM, 2012.

5. Bonaquist, R. F. *NCHRP Report 629: Ruggedness Testing of the Dynamic Modulus and Flow Number Tests with the Simple Performance Tester*. Transportation Research Board of the National Academies, 2008.

6. AASHTO T 378-17. *Determining the Dynamic Modulus and Flow Number for Asphalt Mixtures Using the Asphalt Mixture Performance Tester (AMPT)*. AASHTO, 2017.

7. AASHTO TP 131-18, *Determining the Dynamic Modulus of Asphalt Mixtures Using the Indirect Tension Test*. AASHTO, 2018.

8. Norouzi, A., and Y. R. Kim. Ruggedness Study of Dynamic Modulus Testing of Asphalt Concrete in Indirect Tension Mode. *Journal of Testing and Evaluation*, Vol. 45, No. 2, 2017, pp. 601–612.

9. AASHTO R 83-17. *Preparation of Cylindrical Performance Test Specimens Using the Superpave Gyratory Compactor (SGC)*. AASHTO, 2017.

10. AASHTO PP 99-19. *Preparation of Small Cylindrical Performance Test Specimens Using the Superpave Gyratory Compactor (SGC) or Field Cores*. AASHTO, 2019.

11. AASHTO TP 132-19, *Determining the Dynamic Modulus for Asphalt Mixtures Using Small Specimens in the Asphalt Mixture Performance Tester (AMPT)*. AASHTO, 2019.

12. AASHTO R 84-17, *Developing Dynamic Modulus Master Curves for Asphalt Mixtures Using the Asphalt Mixture Performance Tester (AMPT)*. AASHTO, 2017.

13. Keshavarzi, B. *Prediction of Thermal Cracking in Asphalt Pavements Using Simplified Viscoelastic Continuum Damage (S-VECD) Theory*. Ph.D. dissertation, North Carolina State University, 2019.

14. Olard, F., and H. Di Benedetto. General "2S2P1D" Model and Relation between the Linear Viscoelastic Behaviours of Bituminous Binders and Mixes. *Road Materials and Pavement Design*, Vol. 4, No. 2, 2003, pp. 185–224.

15. North Carolina State University. FlexMAT™ version 2.0. FHWA, 2020.

16. Lee, K., C. Castorena, and Y. R. Kim. Improving the Reliability of Damage Characteristic Curves in the Simplified Viscoelastic Continuum Damage Model. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2672, No. 28, 2017, pp. 493–502.

17. Wang, Y., and Y. R. Kim. Development of a Pseudo Strain Energy-Based Fatigue Failure Criterion for Asphalt Mixtures. *International Journal of Pavement Engineering*, Vol. 20, No. 10, pp. 1182–1192, 2017.

18. AASHTO T 321-17. *Determining the Fatigue Life of Compacted Asphalt Mixtures Subjected to Repeated Flexural Bending*. AASHTO, 2017.

19. Kim, Y. R., Y. Wang, and S. Underwood. *TechBrief: Cyclic Fatigue Index Parameter (S_{app}) for Asphalt Performance Engineered Mixture Design*. FHWA-HIF-19-091, FHWA, 2019.

20. North Carolina State University. FlexPAVE™ version 1.1. FHWA, 2020.

21. Wang, Y. *Development of the Framework of Performance-Engineered Mixture Design for Asphalt Concrete*. Ph.D. dissertation, North Carolina State University, 2019.

22. Plackett, R. L. The Design of Optimum Multifactorial Experiments. *Biometrika,* Vol. 133, 1946, pp. 305–325.

23. Castorena, C., Y.R. Kim, S. Pape, & K. Lee. *Development of Small Specimen Geometries for Asphalt Mixture Performance Testing NCHRP IDEA N181 Final Report.* Transportation Research Board of the National Academies, 2017.

24. Lee, J.S., A. Norouzi, and Y.R. Kim. Determining Specimen Geometry of Cylindrical Specimens for Direct Tension Fatigue Testing of Asphalt Concrete. *Journal of Testing and Evaluation*, Vol. 45, No. 2, 2017, pp. 613–623.

25. AASHTO T 342-11. *Determining Dynamic Modulus of Hot Mix Asphalt (HMA)*. AASHTO, 2011.

26. MATLAB® software version R2018b. The MathWorks Inc., 2018.

27. Luca, J., and D. Mrawira. New Measurement of Thermal Properties of Superpave Asphalt Concrete. *Journal of Materials in Civil Engineering*, Vol. 17, No. 1, 2005, pp. 72–79.

28. Underwood, B. S., A Method to Select General Circulation Models for Pavement Performance Evaluation. *International Journal of Pavement Engineering*, 2019, pp. 1–13.

29. Ang, K. H., G. C. Y. Chong, and Y. Li. PID control system analysis, design, and technology. *IEEE Transactions on Control Systems Technology*, Vol. 13, No. 4, 2005, pp. 559-576.

30. National Cooperative Highway Research Program. *NCHRP Report 629: Ruggedness Testing of the Dynamic Modulus and Flow Number Tests with the Simple Performance Tester*, "Appendix E. Equipment Specification for the Simple Performance Test System," NCHRP, 2008.