NCHRP

REPORT 508

NATIONAL
COOPERATIVE
HIGHWAY
RESEARCH
PROGRAM

Accelerated Laboratory Rutting Tests: Evaluation of the Asphalt Pavement Analyzer

TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

NCHRP REPORT 508

Accelerated Laboratory Rutting Tests: Evaluation of the Asphalt Pavement Analyzer

PRITHVI S. KANDHAL

AND

L. ALLEN COOLEY, JR.
National Center for Asphalt Technology
Auburn, AL

SUBJECT AREAS
Materials and Construction

Research Sponsored by the American Association of State Highway and Transportation Officials
in Cooperation with the Federal Highway Administration

TRANSPORTATION RESEARCH BOARD

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. — 2003

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

Systematic, well-designed research provides the most effective approach to the solution of many problems facing highway administrators and engineers. Often, highway problems are of local interest and can best be studied by highway departments individually or in cooperation with their state universities and others. However, the accelerating growth of highway transportation develops increasingly complex problems of wide interest to highway authorities. These problems are best studied through a coordinated program of cooperative research.

In recognition of these needs, the highway administrators of the American Association of State Highway and Transportation Officials initiated in 1962 an objective national highway research program employing modern scientific techniques. This program is supported on a continuing basis by funds from participating member states of the Association and it receives the full cooperation and support of the Federal Highway Administration, United States Department of Transportation.

The Transportation Research Board of the National Academies was requested by the Association to administer the research program because of the Board's recognized objectivity and understanding of modern research practices. The Board is uniquely suited for this purpose as it maintains an extensive committee structure from which authorities on any highway transportation subject may be drawn; it possesses avenues of communications and cooperation with federal, state and local governmental agencies, universities, and industry; its relationship to the National Research Council is an insurance of objectivity; it maintains a full-time research correlation staff of specialists in highway transportation matters to bring the findings of research directly to those who are in a position to use them.

The program is developed on the basis of research needs identified by chief administrators of the highway and transportation departments and by committees of AASHTO. Each year, specific areas of research needs to be included in the program are proposed to the National Research Council and the Board by the American Association of State Highway and Transportation Officials. Research projects to fulfill these needs are defined by the Board, and qualified research agencies are selected from those that have submitted proposals. Administration and surveillance of research contracts are the responsibilities of the National Research Council and the Transportation Research Board.

The needs for highway research are many, and the National Cooperative Highway Research Program can make significant contributions to the solution of highway transportation problems of mutual concern to many responsible groups. The program, however, is intended to complement rather than to substitute for or duplicate other highway research programs.

NCHRP REPORT 508

Project D9-17 FY'99

ISSN 0077-5614

ISBN 0-309-08775-9

Library of Congress Control Number 2003113650

Price $21.00

NOTICE

The project that is the subject of this report was a part of the National Cooperative Highway Research Program conducted by the Transportation Research Board with the approval of the Governing Board of the National Research Council. Such approval reflects the Governing Board's judgment that the program concerned is of national importance and appropriate with respect to both the purposes and resources of the National Research Council.

The members of the technical committee selected to monitor this project and to review this report were chosen for recognized scholarly competence and with due consideration for the balance of disciplines appropriate to the project. The opinions and conclusions expressed or implied are those of the research agency that performed the research, and, while they have been accepted as appropriate by the technical committee, they are not necessarily those of the Transportation Research Board, the National Research Council, the American Association of State Highway and Transportation Officials, or the Federal Highway Administration, U.S. Department of Transportation.

Each report is reviewed and accepted for publication by the technical committee according to procedures established and monitored by the Transportation Research Board Executive Committee and the Governing Board of the National Research Council.

Published reports of the

NATIONAL COOPERATIVE HIGHWAY RESEARCH PROGRAM

are available from:

Transportation Research Board
Business Office
500 Fifth Street, NW
Washington, DC 20001

and can be ordered through the Internet at:

http://www.national-academies.org/trb/bookstore

Printed in the United States of America

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is a division of the National Research Council, which serves the National Academy of Sciences and the National Academy of Engineering. The Board's mission is to promote innovation and progress in transportation through research. In an objective and interdisciplinary setting, the Board facilitates the sharing of information on transportation practice and policy by researchers and practitioners; stimulates research and offers research management services that promote technical excellence; provides expert advice on transportation policy and programs; and disseminates research results broadly and encourages their implementation. The Board's varied activities annually engage more than 4,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

AUTHOR ACKNOWLEDGMENTS

The research reported herein was performed under NCHRP Project 9-17 by the National Center for Asphalt Technology, Auburn University. Prithvi S. Kandhal, Associate Director, National Center for Asphalt Technology, was the principal investigator. The other author of this report was L. Allen Cooley, Jr., Research Engineer at the National Center for Asphalt Technology. The work was conducted under the general supervision of Messrs. Kandhal and Cooley.

The authors would like to acknowledge contributions to this report by several entities. First, the authors would like to thank APAC Materials Services in Smyrna, Georgia, for the use of their Hamburg Wheel Tracking Device. Next, thanks is due to Pavement Technology, Inc., for the loan of a second Asphalt Pavement Analyzer. We would also like to acknowledge the Federal Highway Administration (FHWA), namely Mr. Kevin Stuart, for identifying an alternate aggregate source for the material placed at the FHWA Accelerated Loading Facility. Finally, the authors appreciate the help of Dr. Sam Carpenter for evaluating the boundary conditions within Asphalt Pavement Analyzer test specimens.

FOREWORD

By Edward T. Harrigan
Staff Officer
Transportation Research Board

This report presents the findings of a research project to determine the suitability of the Asphalt Pavement Analyzer, which is a loaded-wheel tester, (1) as a general method of predicting rutting potential and (2) for use in field quality control and quality acceptance operations. The report will be of particular interest to materials engineers in state highway agencies, as well as to materials suppliers and paving contractor personnel who are responsible for the design and evaluation of hot mix asphalt.

Increased truck traffic, heavier axle loads, higher tire pressures, and increasing use of super-single tires have contributed to the demand for rut-resistant hot mix asphalt (HMA). The prevention of premature rutting of HMA pavements relies on proper mix design, production, and construction. The Strategic Highway Research Program provided several new techniques to prevent rutting failures through laboratory evaluation of HMA materials during the mix design process; however, these tests are often time consuming, require fairly complex test equipment, and are not intended for quality control and quality acceptance (QC/QA). Therefore, work has continued in the public and private sectors on the development of simpler, quicker tests to predict rutting during HMA mix design as well as for QC/QA of HMA production and pavement construction.

The Asphalt Pavement Analyzer (APA) has been widely adopted as a straightforward method to evaluate HMA rutting potential in mix design and QC/QA applications. However, the APA test does not yield a fundamental material property that can be used with appropriate materials characterization and distress prediction models to predict rutting performance. A key issue with the APA (or with any other method intended for this purpose, including other types of loaded wheel testers and simple strength tests) is the degree to which the relationship between the APA's test results and actual field performance depends on specific project-associated factors such as aggregate properties, mix design type, traffic level, and traffic speed. The utility of the APA will be enhanced if it is known to provide results that are directly comparable across disparate projects.

Under NCHRP Project 9-17, "Accelerated Laboratory Rutting Tests: Asphalt Pavement Analyzer," the National Center for Asphalt Technology (NCAT) at Auburn University was assigned the tasks of (1) evaluating the APA to determine its suitability as a general method of predicting rutting potential and for use in field QC/QA operations and (2) comparing its effectiveness with that of other loaded wheel testers and simple performance tests such as those identified in NCHRP Project 9-19, "Superpave Support and Performance Models Management."

The research team reviewed the literature on the development and use of the APA and used the results of the review to design and carry out a comprehensive program of laboratory testing with HMA mixes of known performance to validate the APA as a general method of predicting rutting potential. The laboratory testing program was organized in two phases and made extensive use of original materials, loose mix, and performance data from several large-scale accelerated pavement tests—WesTrack, MnRoad, the FHWA Accelerated Loading Facility, and the NCAT Test Track—and from the Nevada DOT I-80 field experiment. In the first phase, numerous APA testing conditions were evaluated to develop a preliminary APA test method whose results showed the best relationship to measured field performance. In the second phase, this preliminary method was validated with results obtained from a second, independent set of HMA materials and field performance data.

Through the first phase of the testing program, the research team found that the APA test method with the best relationship to performance featured the use of (1) cylindrical specimens compacted to 4-percent air voids or beam specimens compacted to 5-percent air voids, (2) a test temperature corresponding to the high temperature of the standard binder performance grade for the project location, and (3) a standard APA linear hose.

The validation phase of the testing program established that laboratory rut depths measured with the preliminary APA method correlated well with field performance on an individual project basis. However, it was not generally possible to predict field rut depths from APA testing on any given project using relationships developed from other projects with different geographic locations and traffic.

Further, the report identified several issues that will need to be addressed through future research before the issue of the suitability of the APA as a field QC/QA method can be settled. Finally, the report presents limited relationships between field rut depths and test results comparing the APA with the Hamburg Wheel Tracking Device and the simple performance tests from NCHRP Project 9-19.

This final report includes a detailed description of the experimental plan, a discussion of the research results, and two supporting appendixes:

- Appendix A: Literature Review; and
- Appendix B: Preliminary APA Test Method.

The research results have been referred to the TRB Mixtures and Aggregate Expert Task Group for its review and possible recommendation to the AASHTO Highway Subcommittee on Materials for adoption of the draft test method presented in Appendix B.

CONTENTS

ACCELERATED LABORATORY RUTTING TESTS: EVALUATION OF THE ASPHALT PAVEMENT ANALYZER

SUMMARY The conclusions from this study are based upon the research results obtained during two phases of research. The first phase of work was conducted to identify test conditions within the Asphalt Pavement Analyzer (APA) that produced results most related to field rutting performance. The results were used to recommend a tentative APA test method. Based on this phase of work, cylindrical samples compacted to 4-percent air voids and beam samples compacted to 5-percent air voids resulted in APA laboratory test results that were more closely related to field rutting performance than were cylindrical and beam samples compacted to 7-percent air voids. Samples tested in the APA at a test temperature corresponding to the high temperature of the standard performance grade for a project location better predicted field rutting performance than did samples tested at 6°C higher than the high temperature of the standard performance grade. Samples tested with both the standard and large-diameter hoses predicted field rutting performance about equally; however, samples tested with the standard hose produced less variability. Beam and cylindrical samples predicted field rutting performance about equally.

The second phase of the study was conducted to validate the proposed APA test method. Laboratory rut depths measured by the APA had good correlations on individual project basis; however, it is generally not possible to predict field rut depths from APA testing on a specific project using relationships developed from other projects with different geographical locations and traffic.

CHAPTER 1

INTRODUCTION AND RESEARCH APPROACH

1.1 INTRODUCTION

Permanent deformation, or rutting, has been and continues to be a problem in the performance of hot mix asphalt (HMA) pavements. Rutting is defined as the accumulation of small amounts of unrecoverable strain resulting from applied loads to the pavement. This deformation is caused by the consolidation, a lateral movement of the HMA under traffic, or both. Shear failure (lateral movement) of the HMA courses generally occurs in the top 100 mm of the pavement surface (*1*). Rutting not only decreases the useful service life of the pavement, but also creates a safety hazard for the traveling public. In recent years, the potential for rutting on the nation's highways has increased due to higher traffic volumes (equivalent single axle loads [ESALs]) and the increased use of radial tires, which typically exhibit higher inflation pressures.

A standardized accelerated laboratory test to predict HMA rutting potential that is relatively inexpensive and useful for quality control/quality assurance (QC/QA) testing would be of great benefit. Currently the most common type of standardized laboratory test of this nature is a loaded wheel tester (LWT). Numerous types of LWT equipment are available, such as the Georgia Loaded Wheel Tester, the Asphalt Pavement Analyzer (APA), the Superfos Construction Rut Tester, the Hamburg Wheel Tracking Device, and the French Laboratoire Central des Ponts et Chaussées (LCPC) Wheel Tracker.

In an effort to identify HMA mixtures that may be prone to rutting, many agencies have begun using LWTs as supplements to their mixture design procedures. The LWTs allow accelerated proof testing of mix designs.

In order for LWT devices to be used with a significant level of confidence, there needs to be an acceptable correlation of actual field rutting to those values predicted by LWTs in the laboratory. Some of the agencies using LWTs have recognized this fact and have conducted research projects to determine the degree of correlation between field performance and laboratory LWTs.

Because of the successes of some agencies using the APA, the objectives of project, NCHRP Project 9-17, are as follows:

1. Evaluate the APA to determine its suitability as a general method of predicting rut potential and for use in field QC/QA testing and
2. Compare the effectiveness of the APA with that of other LWTs and with a simple strength test.

This project will focus on the APA because it is the LWT most widely used in the United States. Where possible, APA results will be compared with results obtained on a common set of materials from other LWTs in order to provide a link to past results with the other devices and to better estimate their effectiveness compared with the APA.

1.2 DESCRIPTION OF THE APA

The original version of the APA was the Georgia Loaded Wheel Tester (GLWT), shown in Figure 1, which was developed during the mid-1980s through a cooperative research study between the Georgia DOT and the Georgia Institute of Technology (*2*). Development of the GLWT consisted of modifying a wheel-tracking device originally designed by C.R. Benedict of Benedict Slurry Seals, Inc., to test slurry seals (*2*). The primary purpose for developing the GLWT was to perform efficient, effective, and routine laboratory rut proof testing and field production quality control of HMA (*2*).

The APA, shown in Figure 2, is a modification of the GLWT and was first manufactured in 1996 by Pavement Technology, Inc. The APA has been used to evaluate the rutting, fatigue, and moisture resistance of HMA mixtures. Because the APA is the second generation of the GLWT, it follows a similar rut-testing procedure. A wheel is loaded onto a pressurized linear hose and tracked back and forth over a testing sample to induce rutting.

1.3 SCOPE

To accomplish the research objectives, eight tasks were conducted; their descriptions follow.

1.3.1 Phase I

Task 1: Determine the State of Practice

In Task 1, a literature search and review was conducted to determine the current state of practice for the APA as

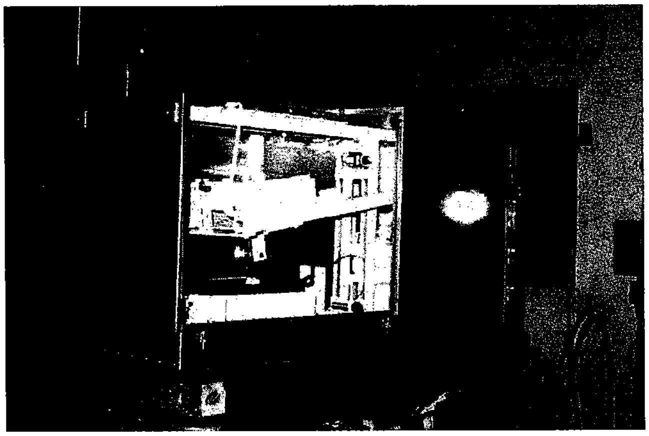


Figure 1. Georgia Loaded Wheel Tester.

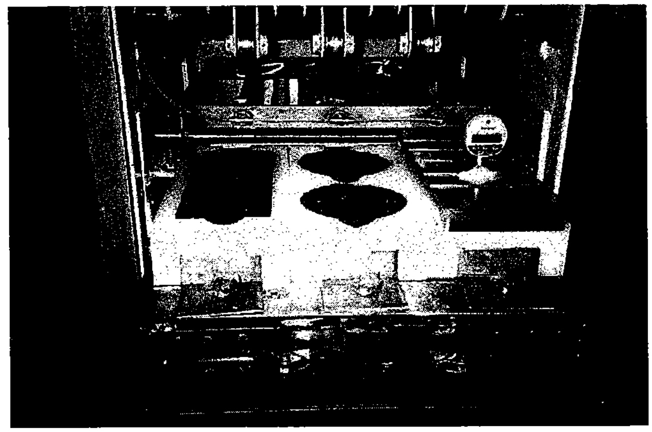


Figure 2. Asphalt Pavement Analyzer.

well as for other LWTs. Results of Task 1 are presented in Chapter 2.

Task 2: Develop Experimental Plan

In order to evaluate the ability of the APA to predict field rutting potential, a statistically controlled experimental plan was developed. A significant effort was applied to identify HMA mixes with known field performance that represented a range of climatic, materials, and project characteristics. Those HMA mixtures are described in Chapter 3. Chapter 4 details the experimental plan.

Task 3: Submit Interim Report

Task 3 involved preparing an Interim Report documenting the results of Tasks 1 and 2. The Interim Report was submitted to NCHRP in August 1999.

1.3.2 Phase II

Task 4: Conduct Laboratory Work

After approval of the experimental plan contained in the Interim Report, testing of the approved HMA mixes was conducted. A summary and analysis of those results are provided in Chapter 5.

Task 5: Develop Preliminary Test Method in AASHTO Format

Based upon the conclusions drawn from Task 4, a preliminary AASHTO test method for the APA was developed. This method is provided in Appendix B.

Task 6: Submit Interim Report

In December of 2000, an Interim Report was submitted to document the research conducted in Tasks 1 through 5. Also included in this report was an experimental plan to validate the preliminary test method. Fourteen additional mixes of known field performance were identified for this experiment. Chapter 6 details the work plan and mixes for Task 6.

1.3.3 Phase III

Task 7: Validate Preliminary Test Method

In Task 7, the 14 new HMA mixes were tested and analyzed to validate the proposed test method. The results of this analysis are presented in Chapter 7.

Task 8: Final Report

This final report was prepared documenting all of the research conducted in Tasks 1 through 7. Chapters 8 through 10 present (1) a future work plan for an extended field validation of the APA method, (2) a recommended practice for tailoring the APA specification criteria to local conditions, and (3) a summary of the conclusions of NCHRP Project 9-17, respectively.

CHAPTER 2
STATE OF THE PRACTICE (PHASE I)

2.1 INTRODUCTION

Task 1 of the project reviewed the literature for information on the state of the practice for the APA and other LWTs. This review was to address each LWT's suitability for predicting the rutting potential of HMA during laboratory mix design and QC/QA testing.

The literature review was conducted to specifically answer the following questions:

1. What are the key test parameters, limitations, material sensitivities, and boundary conditions utilized by various LWTs?
2. What are the conclusions and recommendations of researchers who have evaluated various LWTs, specifically the suitability of LWTs to predict rutting?
3. What areas need further evaluation and standardization to verify or improve the APA's ability to predict rutting?

The predominant LWTs found in the literature were the French, Hamburg, and APA LWTs and the GLWT. However, some references were found that included other types of LWTs. Additionally, some publications compared different LWTs during research. Detailed summaries of articles and papers found in the literature are provided in Appendix A. Following is a summary of the state of the practice for the APA. Also described are potential factors or variables that were evaluated or standardized during NCHRP Project 9-17. For those variables in which the literature suggests no further evaluation or standardization is needed, justification is provided.

The APA is essentially the second generation of the GLWT. The GLWT was developed during the mid-1980s through a cooperative research study between the GDOT and the Georgia Institute of Technology (GIT) (*2*). The GIT modified an existing machine developed by C.R. Benedict of Benedict Slurry Seals, Inc., for designing and testing slurry seals (*2*). Subsequent research studies conducted by GIT for GDOT made further improvements to the GLWT (*3*, *4*). During the mid- to late-1990s, Pavement Technology, Inc., developed the APA. This LWT combined the testing methodologies of the GLWT with some improvements; therefore, it has been called the second generation of the GLWT.

During the literature review, numerous variables were identified for potential inclusion in the experimental design for NCHRP Project 9-17. Table 1 presents these variables and different levels for each factor.

Obviously, the time and budget constraints for this project did not allow for all factor-level combinations listed in Table 1 to be studied; however, many of the variables have been studied by other researchers. The following paragraphs discuss each of the variables and identify whether additional evaluation was warranted.

The first factor in Table 1 is compaction method for samples. The original compactor used by Lai was a kneading compactor (*2*) with a "loaded" foot. Subsequent compactors used with the GLWT and APA were a compression machine (*5*, *6*); a rolling compaction machine (*7*, *8*, *9*); a Superpave® Gyratory Compactor (SGC) (*9*, *10*, *11*, *12*, *13*); an Asphalt Vibratory Compactor (AVC) (*11*, *13*, *14*, *15*); and a vibrating hammer (*16*, *17*). In the research studies conducted within the last few years, the SGC and AVC have been the predominant compactors used. Several studies have compared cylindrical specimens from the SGC and beam specimens from the AVC. Collins et al. (*10*) suggested that the two specimen types ranked the mixtures similarly; however, it was shown that they do not provide similar rut depths. Because these two types of compactor rank mixtures similarly but provide significantly different rut depths, they were included within the experimental plan.

The second factor in Table 1 is test temperature. The original work by Lai used a test temperature of 35°C (95°F) (*2*). This temperature was selected because it corresponded to the average high summer air temperature in Georgia. Since Lai's original work, test temperatures have steadily increased for the GLWT and APA. The most recent work has stated (or inferred) that testing should be conducted either at expected high pavement temperatures or at the high temperature of the standard performance grade—that is, at the performance grade of an asphalt binder before any "bumping." For instance, based upon the program LTPP Bind, the standard performance grade for Auburn, Alabama, is a PG 64-16 at a 98-percent reliability. The standard performance grade for Minneapolis, Minnesota, is a PG 58-34, also at a 98-percent reliability. Therefore, using the argument that testing in the APA should be conducted at the high temperature of the standard performance grade, mixes in Auburn would be tested at 64°C while mixes in Minneapolis would be tested at 58°C. LTPP Bind provides actual pavement temperatures of 55.0 and 46.7°C at

TABLE 1 Compilation of variables for LWT devices determined from literature review

Factor	Level
Compaction Method for Samples	Rolling Wheel Compactor
	Superpave Gyratory Compactor
	Static Compression Machine
	French Plate Compactor
	Linear Kneading Compactor
	Slabs/Cores Cut From Existing Pavement
	Vibratory Tamper
	Vibratory Compactor (Pavetec)
Test Temperature	35°C
	40°C
	40.6°C
	46.1°C
	50°C
	55°C
	58°C
	60°C
	64°C
Sample Conditioning	24 h at test temperature
	Submerged in 50°C water bath
	6 h at test temperature
	AASHTO T283 freeze-thaw cycle
Aging of Loose Mixture	2 h at 135°C
	4 h at 135°C
Terminal Number of Cycles	1,000
	2,000
	4,000
	8,000
	10,000
	20,000
	30,000
Air Void Content	Some Standard Compactive Effort
	4 ± 1 percent
	6 percent
	7 percent
	7 ± 1 percent
	8 ± 1 percent
Loading Apparatus	Rubber Hose
	Solid Steel Wheel
	Pneumatic Tire
	Hard Rubber Tire
Nominal Maximum Aggregate Size	9.5 mm
	12.5 mm
	19.0 mm
	25.0 mm
	37.5 mm
Asphalt Binder	AC-5
	AC-10
	AC-20
	AC-30
	PG 58-34
	PG 58-22
	PG 64-22
	PG 76-22
	PG82-22
Test Specimen Size (l x w x thickness) or (diameter x thickness)	381 mm x 76 mm x 76 mm
	300 mm x 125 mm x 75 mm
	320 mm x 120 mm x 80 mm
	500 mm x 180 mm x 100 mm
	150 mm x 75 mm
"Linear" Hose or Pneumatic Tire Pressure	517 kPa
	600 kPa
	690 kPa
	827 kPa
	900 kPa
Wheel Load	222 N
	334 N
	445 N
	543 N
	660 N
	700 N
	710 N
	5,000 N
Loading Direction	Cyclic (back and forth)
	Uni-Directional (one direction)
Rate of Loading	Different LWTs Use Different Rates
Type of Base Plate	Rigid Base Plate (simulate PCC)
	Flexible Base Plate (simulate HMA)
Hose Stiffness	No Specifics Provided
Gradation	Varying
Asphalt Content	Varying

50-percent reliability for Auburn and Minneapolis, respectively. It is interesting to note that Williams and Prowell (*18*) conducted testing on WesTrack mixes with the APA at 60°C and achieved a correlation coefficient (R^2) value of 89.9 when comparing APA results with field results. The 98-percent reliability high temperature for the performance grade at the Reno, Nevada, Cannon International Airport (which is not a great distance from WesTrack) is 64°C (PG 64-22). The difference between these two temperatures is only 4°C. Therefore, the experimental plan included a test temperature corresponding to the standard high temperature of the performance grade for a project location. This approach is practical (and conservative) and easy to implement if the APA is used by the highway agencies.

The next variable in Table 1 is sample conditioning. With the APA, sample conditioning refers to the time a test sample must be heated at the prescribed test temperature prior to testing. West (*13*) conducted the only reported formal study that has evaluated preheat time. According to his work, 6 and 24 h of preheat time did not provide statistically different rut depths. However, West used 55 and 60°C. If the standard high temperature of the performance grade is used as the test temperature, the highest possible test temperature in the United States is 70°C. Because previous research has indicated the significance of test temperature, preheat time was a candidate for further investigation.

Aging of loose mixture in the laboratory is the next factor in Table 1. Only two protocols are currently available for aging HMA mixtures in the laboratory: the Superpave short- and long-term procedures (AASHTO PP2). Recommendations from NCHRP Project 9-9 (which are published in *NCHRP Research Results Digest 237* and *CRP-CD-1*) for short-term aging were to age loose mixture for 2 h at compaction temperature if aggregates within the mixture had water absorption values less than 2.0 percent. Stuart and Izzo (*16*) suggested 2 h at 135°C. For most neat asphalt binders, compaction temperature is not much different than 135°C. Therefore, within NCHRP Project 9-17, loose mixtures were aged for 2 h at compaction temperature in order to comply with Superpave procedures.

The next factor in Table 1 is terminal number of cycles. Within the literature, the standard terminal number of cycles has been 8,000. Lai conducted some testing to 10,000 cycles (*2*); however, some studies have suggested that a lower number of cycles may be sufficient (*5, 8*). In order to thoroughly investigate the APA, it was decided to conduct testing to 10,000 cycles with rut depths obtained at 1,000 cycle increments.

Specimen air void content is the next factor in Table 1. The literature indicates that for a given mixture and test temperature, as the air void content increases, so do measured rut depths. However, the question that may be more pertinent is what air void content should be used in the laboratory to obtain the best correlation with field performance. Some researchers believe that the air voids should be around 7 percent because this air void content generally represents normal construction as well as specified values. To best simulate permanent deformation in the field, laboratory rut depths should include deformation that occurs due to the consolidation by the action of traffic. Some researchers believe that 4-percent air voids (also in design) should be used because actual shear failure of a mixture usually takes place below 4 percent. Therefore, to resolve this issue, testing in NCHRP Project 9-17 was conducted at both 7 and 4 percent air voids. Based on the work conducted by West (*13*), a tolerance of ±0.5 percent on air voids was standardized for this study. Since obtaining 4-percent air voids is generally difficult for beam specimens compacted with AVC, a value higher than 4 percent (5 percent) was used.

The next variable presented in Table 1 is loading apparatus. The levels presented for this factor correspond to the different LWTs found in the literature. A linear rubber hose is a standard piece of the APA equipment. Therefore no further evaluation was conducted.

Nominal maximum aggregate size (NMAS) is the next factor. Several studies indicated that mixes with larger NMAS values typically have lower rut depths (*3, 9, 12, 17*). This factor will indirectly be evaluated within NCHRP Project 9-17 as 10 different mixes from various locations throughout the United States will be evaluated, as noted in the experimental plan (Chapter 4). The next factor in Table 1, asphalt binder type, was also indirectly evaluated because the 10 mixtures used a different binder. The next factor, test specimen size, was evaluated in terms of the use of slab and cylindrical specimens.

Pressure within the linear hose is the next factor in Table 1. This factor, along with the next factor, wheel load, has been very consistent since Lai's original work (*2, 3,* and *4*). A hose pressure of 690 kPa (100 psi) was predominant in the literature. Similarly, wheel load has also been fairly consistent at 445 N (100 lb). However, Williams and Prowell (*18*) used a hose pressure/wheel load combination of 830 kPa/533 N (120 psi/120 lb) and found a strong correlation between laboratory and field rutting. Based upon work by Wu (*19*), the contact pressure and contact area for a hose pressure/wheel load combination of 690 kPa/445 N were 689 kPa and 6.45 cm^2, respectively. The contact pressure and contact area for 830 kPa/534 N were 730 kPa and 7.29 cm^2. It was desirable to further investigate the two hose pressure/wheel load combinations (690 kPa/445 N and 830 kPa/534 N).

The next factor in Table 1 is loading direction. This factor could potentially affect results within the APA; however, the cost of such a factor could be significant as the APA would have to be redesigned and modified. Therefore, loading direction was not included within NCHRP Project 9-17.

Rate of loading is another factor for which no specific research has been conducted. The research agency included this factor in Table 1 because it was felt that the rate of loading may amplify the differences between good and bad performing mixes. This factor should be investigated in future work with the APA.

The next factor in Table 1 is type of base plate. This is another factor for which no literature was found. A base plate that is flexible and has the same modulus as HMA could be placed under samples of HMA intended for placement over an existing HMA pavement. Likewise, a rigid plate could be placed under samples of HMA intended for placement over an existing Portland cement concrete (PCC) pavement. The effect of base plate type could affect rut depths and may more accurately simulate conditions in the field. However, since no research has been conducted on this complex subject, it was not studied in NCHRP Project 9-17 because of budget and time constraints.

The next factor in Table 1 is hose stiffness. Lai (*20*) did evaluate hose stiffness, but no specifics of the actual stiffness were provided. Another possibility for researching this type of factor would be to evaluate different hose diameters. A factor of this nature would also affect the contact area that a linear hose has on a sample. A hose diameter larger than the 25 mm currently used in the APA may be desirable, especially for large stone mixes.

The final two factors are gradation and asphalt content, which were indirectly evaluated through the use of a variety of experimental mixes.

CHAPTER 3

SELECTION OF MATERIALS (PHASE I)

3.1 10 HMA MIXES OF KNOWN PERFORMANCE

In accordance with the approved work plan, 10 HMA mixes of known field rutting performance were used in the Task 4 experiment. The following sections describe the materials selected.

The approved work plan required the use of 10 HMA mixes of known rutting performance within a full factorial experiment designed to determine the combination of testing conditions for the APA that best predicts field rutting. These 10 mixes were selected from three full-scale pavement research projects and encompass climatic regions, project characteristics, and materials from throughout the United States: WesTrack (Nevada), the Minnesota Road Research Project (MnRoad), and the FHWA Accelerated Loading Facility (ALF) at Turner-Fairbank Highway Research Center (Virginia). The following sections describe the mixes selected from each of these projects.

3.1.1 Mixes from WesTrack

WesTrack was a federally funded accelerated full-scale HMA pavement research project to develop performance-related specifications for HMA and also to provide early field verification for the Superpave asphalt mixture design procedure.

The original WesTrack experiment consisted of 26 different HMA test sections located on a 3-km (1.8-mile) oval test track 60 miles east of Reno, Nevada. Loading was achieved using four driverless tractor trailers. Loads on each axle were 80 kN (20,000 lb), contributing to a total of 10.3 ESALs per truck pass. The trucks were fitted with 295/75R22.5 radial tires inflated to 690 kPa (100 psi). The test speed was 64 km/h (40 mph).

For the original 26 sections, a single Superpave performance-graded asphalt binder (PG 64-22) was used. A single primary source of aggregate was selected for use in the test sections. Three different 19-mm aggregate gradations were used. One gradation was located on the fine side of the restricted zone and designated as "fine." The second gradation also passed on the fine side of the restricted zone; however, additional baghouse fines were added to the gradation during production. This gradation was designated "fine-plus." The third gradation passed on the coarse side of the restricted zone and was designated "coarse."

Using the Superpave volumetric mix design procedure, optimum asphalt contents were determined for the coarse and fine gradations. Optimum asphalt content determined for the fine gradation was also used for the fine-plus gradation. During construction, asphalt contents for all three gradation types were varied by ± 0.7 percent. Additionally during construction, three levels of in-place air voids were specified. An "optimum" in-place air void content was chosen as 8 percent, while variations of ± 4.0 percent were also planned.

For NCHRP Project 9-17, test sections were selected from WesTrack that were placed at "optimum" construction. This infers that sections were selected at optimum asphalt content and at optimum in-place air void contents. Additionally, a criterion for selecting test sections for inclusion in NCHRP Project 9-17 was that all three gradations be included, which would allow a comparison of rut depths among different gradations when the same aggregate and binder source are used. Therefore, WesTrack Test Sections 15 (fine gradation), 19 (fine-plus gradation), and 24 (coarse gradation) were selected. Of these three sections, two (Sections 15 and 24) were also used in NCHRP Project 9-19, "Superpave Support and Performance Models Management," for development and validation of the simple performance test (see *NCHRP Research Results Digest 237* and *CRP-CD-1).*

Table 2 provides design and in-place gradations and asphalt contents for these three sections. Gradations and asphalt contents for Section 19 are based upon average values obtained from cores. In-place data for Sections 15 and 24 presented in the table were provided by NCHRP Project 9-19 and were obtained from quality control testing and cores. The in-place data indicate that Sections 15 and 24 were placed at filler contents (percent passing 0.075-mm sieve) that were slightly lower than the design values. Section 24 was placed at a 0.2-percent higher asphalt content.

Each of the selected test sections exhibited different performance in the field with respect to rutting. Sections 15 and 19 had total rut depths of 9.2 and 14.5 mm, respectively, after 5 million ESALs; Section 24 failed with 23.0 mm of total rut depth at 2.8 million ESALs. Figure 3 illustrates the downward rut depths versus ESALs for these three sections.

TABLE 2 Mixture information for test sections from WesTrack (design and in-place)

DESIGN VALUES				
Sieve Size		Test Sections (Gradation)		
U.S.	**mm**	**15 (Fine)**	**19 (Fine-Plus)**	**24 (Coarse)**
¾ in.	19.0	100	100	100
½ in.	12.5	91	91	85
_ in.	9.5	79	78	67
No. 4	4.75	49	50	41
No. 8	2.36	38	39	28
No. 16	1.18	34	35	21
No. 30	0.06	28	29	16
No. 50	0.03	17	17	13
No. 100	0.150	8	9	10
No. 200	0.075	5.5	5.7	7.5
Asphalt Content, %		*5.4*	*5.4*	*5.7*
IN-PLACE VALUES				
U.S.	**mm**	**15 (Fine)**	**19 (Fine-Plus)**	**24 (Coarse)**
¾ in.	19.0	100.0	100.0	100.0
½ in.	12.5	87.8	86.2	80.8
_ in.	9.5	75.9	75.1	65.8
No. 4	4.75	49.4	51.4	42.0
No. 8	2.36	38.0	40.3	28.2
No. 16	1.18	33.6	35.8	20.6
No. 30	0.60	27.4	29.5	15.6
No. 50	0.30	15.6	17.9	11.6
No. 100	0.150	7.8	9.3	8.5
No. 200	0.075	4.7	5.8	6.1
Asphalt Content, %		*5.55*	*5.41*	*5.91*

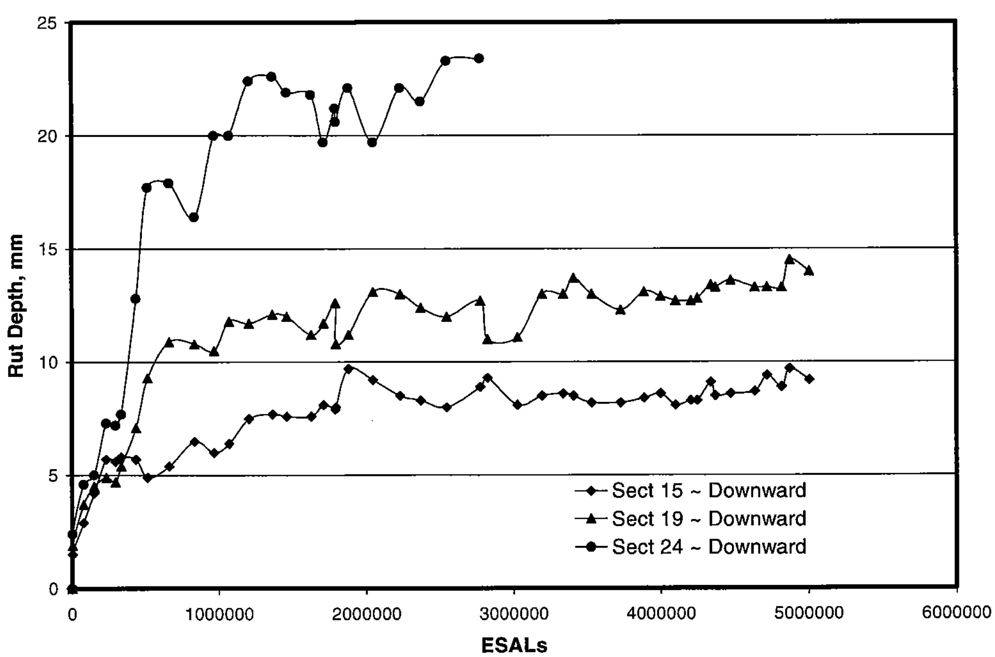


Figure 3. Accumulated downward rut depths for selected WesTrack sections.

3.1.2 Mixes from MnRoad

Three mixes were selected from the MnRoad full-scale pavement study. The MnRoad facility runs parallel to Interstate 94 near Otsego, Minnesota. Traffic from Interstate 94 is periodically diverted onto the facility in order to provide "live" loads. Test sections were selected from the MnRoad mainline sections with a 10-year design life.

Test sections within MnRoad are referred to as cells. The three cells selected for this study were Cell 16, Cell 20, and Cell 21; these contain identical HMA pavement structures. A 38-mm (1.5-in.) wearing course overlies 160 mm (6.3 in.) of HMA base material. The only difference between the pavement structures is that Cell 21 has 580 mm (23 in.) of crushed aggregate base while Cells 16 and 20 have 711 mm (28 in.) of crushed aggregate base. Cells 16 and 20 have also been included within the NCHRP Project 9-19 experiment.

Design and constructed gradations for the three mixes were similar (see Table 3) and had a 12.5-mm NMAS. The primary differences among the three cells was the type of asphalt binder and the method of mix design used to determine optimum asphalt content. Cell 16 used an AC-20; Cells 20 and 21 used an 120/150 penetration graded asphalt binder. Cell 16 was designed using an SGC (design number of gyrations [N_{design}] = 100) while Cells 20 and 21 were designed with a Marshall hammer using 35 and 50 blows, respectively.

Field rut depths for each cell are presented in Table 4. Data for Cells 16 and 21 were obtained in April 2000. Cell 20 data were obtained in April 1999. As seen in Table 4, Cell 20 had significantly more rutting than did Cells 16 and 21, Cell 21 had a moderate amount of rutting, and Cell 16 had the least. Development of field rut depths versus applied ESALs is illustrated in Figure 4. This figure suggests that the Cell 20 mix had begun tertiary flow since the rut depth begins to increase sharply at approximately 2,000,000 ESALs; also, Cell 21 has steadily increasing rut depth over the life of the pavement. Both Cells 20 and 21 appear to be rutting-prone mixes.

TABLE 3 Mixture information for selected cells from MnRoad (design and in-place)

DESIGN VALUES				
Sieve Size		Cell Number		
U.S.	mm	16	20	21
¾ in.	19.0	100	100	100
½ in.	12.5	92	92	92
_ in.	9.5	82	82	82
No. 4	4.75	67	67	67
No. 8	2.36	58	58	58
No. 16	1.18	48	48	48
No. 30	0.06	36	36	36
No. 50	0.03	19	19	19
No. 100	0.150	6	6	6
No. 200	0.075	3.9	3.9	3.9
Asphalt Content, %		*5.6*	*6.4*	*6.1*
IN-PLACE VALUES				
U.S.	mm	16	20	21
¾ in.	19.0	100.0	100.0	100.0
½ in.	12.5	93.4	93.4	93.4
_ in.	9.5	84.7	84.7	84.7
No. 4	4.75	68.8	68.8	68.8
No. 8	2.36	59.5	59.5	59.5
No. 16	1.18	48.2	48.2	48.2
No. 30	0.60	32.9	32.9	32.9
No. 50	0.30	19.5	19.5	19.5
No. 100	0.150	6.7	6.7	6.7
No. 200	0.075	4.7	4.7	4.7
Asphalt Content, %		*5.1*	*6.1*	*5.9*

TABLE 4 Field rut depth data for MnRoad cells

Rut Depth Data	Cell No.		
	16	**20***	**21**
Total Rut Depth, mm	5.3	18.8	12.1
ESALs	3,051,267	2,423,667	3,051,267

*Cell 20 was rehabilitated in April 1999. Reported rut depth and ESALs were in April 1999, before rehabilitation.

3.1.3 Mixes from Turner-Fairbank Highway Research Center

Four mixes were selected from the ALF experiment at Turner-Fairbank Highway Research Center (TFHRC). The ALF experiment consisted of 12 lanes, with four test sites in each lane. The ALF is a structural frame that is 29 m (95 ft) in length and contains a moving wheel assembly designed to model one-half of a single rear truck axle. During ALF testing of mixes selected for NCHRP Project 9-17, a super single tire was used on the wheel assembly. This tire was inflated to a pressure of 690 kPa (100 psi) and was loaded to 43 kN (9667 lb). The tire tracked across the experimental test pavements at 18.5 km/h (11.5 mph) with no wheel wander. An infrared heating system was used to maintain a given test temperature in the pavement.

The four mixes selected for this study were obtained from Lanes 5, 10, 7, and 12. All four of these mixtures were also selected for use in NCHRP Project 9-19. Mixes from Lanes 5, 7, and 10 had identical 19.0-mm NMAS gradations; the Lane 12 mix was a 37.5-mm NMAS. Table 5 presents the design and in-place mixture information for various NMAS gradations, including the 19.0-mm and 37.5-mm NMAS gradations, used for the ALF. In-place values were obtained from researchers for NCHRP Project 9-19.

Three different asphalt binders were used in the four selected mixes. Lanes 10 and 12 both used an AC-20, while Lane 5 had an AC-10 and Lane 7 used a polymer (styrene-butadiene-styrene [SBS] polymer-modified binder) modified binder. Testing of these binders by FHWA indicated that the AC-10 was a PG 58-28, the AC-20 a PG 64-22, and the polymer-modified binder a PG 82-22. Optimum asphalt content for Lanes 5, 7, and 10 was 4.9 percent, while Lane 12 had an optimum binder content of 4.1 percent. Mixing and compaction temperatures for these three binders were as follows:

	AC-10	*AC-20*	*SBS polymer-modified binder*
Mixing:	146°C	154°C	173°C
Compaction:	136°C	143°C	158°C

Rutting performance for each of the four mixes is presented in Table 6. This table shows that the polymer-modified mix (Lane 7) produced less rutting for a given number of wheel

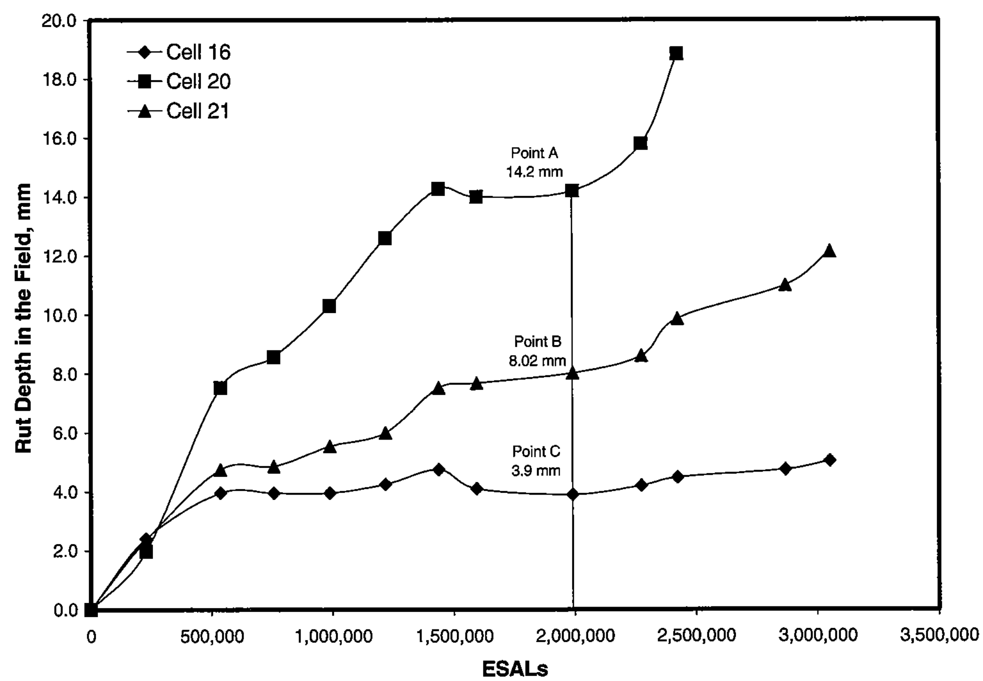


Figure 4. Rut depth versus ESALs for MnRoad.

TABLE 5 Mixture information for selected ALF lanes (design and in-place)

DESIGN VALUES			
Sieve Size		Lanes	
U.S.	mm	Lanes 5, 7, and 10	Lane 12
1½ in.	37.5	100	100
1 in.	25.0	100	86
¾ in.	19.0	99	74
½ in.	12.5	76	65
_ in.	9.5	62	59
No. 4	4.75	44	48
No. 8	2.36	32	32
No. 16	1.18	24	24
No. 30	0.60	17	17
No. 50	0.30	11	12
No. 100	0.150	8	8
No. 200	0.075	5.1	5.7
Asphalt Content, %		*4.9*	*4.0*
IN-PLACE VALUES			
U.S.	**mm**	**Lanes 5, 7, and 10**	**Lane 12**
1½ in.	37.5	100	100
1 in.	25.0	100	85.6
¾ in.	19.0	98.7	73.9
½ in.	12.5	76.0	65.1
_ in.	9.5	62.0	59.0
No. 4	4.75	44.0	47.6
No. 8	2.36	32.5	32.5
No. 16	1.18	23.5	24.0
No. 30	0.60	17.5	17.4
No. 50	0.30	11.5	12.3
No. 100	0.150	8.0	8.0
No. 200	0.075	5.1	5.7
Asphalt Content, %		*4.8, 4,9, 4.9**	*4.1*

*Respectively, Lane 5, Lane 7, and Lane 10.

TABLE 6 Rut performance for ALF mixes

Number of ALF Passes	Rut Depth, mm			
	Lane 5, Site 2	*Lane 10, Site 2*	*Lane 7, Site 2*	*Lane 12, Site 1*
0	0.0	0.0	0.0	0.0
1,000	14.7	15.5	6.3	10.6
4,000	27.4	*	*	*
5,000	*	27.4	7.1	14.4
10,000	*	36.3	12.0	15.2
25,000	*	*	14.3	18.2
75,000	*	*	17.0	21.2
200,000	*	*	18.1	24.1

*Data not available.

passes. The mix containing AC-10 asphalt binder (Lane 5) had significantly more rutting than the other three mixes. Testing in the ALF was conducted at 58°C for all four mixes. Accumulation of rut depths versus the number of wheel passes are presented in Figure 5. This figure shows that Lanes 5 and 10 exhibited similar rutting characteristics. Lanes 7 and 12 were more resistant to permanent deformation than were Lanes 5 and 10 and had somewhat similar rutting characteristics.

One of the problems associated with using mixtures from the ALF was that loadings were characterized by wheel passes instead of by ESALs. In conversations with the research team at TFHRC, little work had been done to convert ALF passes to ESALs. For the analyses included in this report, it was important to convert wheel passes to ESALs because traffic level must be taken into account. The only literature found in which a conversion was attempted was published by Aurilio et al. (*21*). Within this paper, the authors presented a method for converting ALF wheel passes to ESALs. The method entailed assuming a 20-year design traffic level; an estimated traffic volume of 10,000,000 ESALs; and the number of days the pavement temperature at a depth of 20 mm was equal to or higher than 50°C as determined at the weather station nearest the ALF facility at McLean, Virginia. From this information, the authors determined a rutting rate (in mm/year) for the design life and traffic. Results of this analysis are presented in Table 7. Obviously, these results are not realistic for Lanes 5 and 7 because the rut depths were many times higher than the lift thickness of the mixes.

Loaded wheel testing was conducted by FHWA on mixtures from each of the lanes selected for NCHRP Project 9-17. This testing was conducted with the French LCPC Tester, the GLWT, and the Hamburg Wheel Tracking Device (*22*). Results from these LWTs are presented in Table 8. Results from all three test methods indicate that Lane 7 had the least potential for rutting, followed by Lane 12, Lane 10, and Lane 5, respectively.

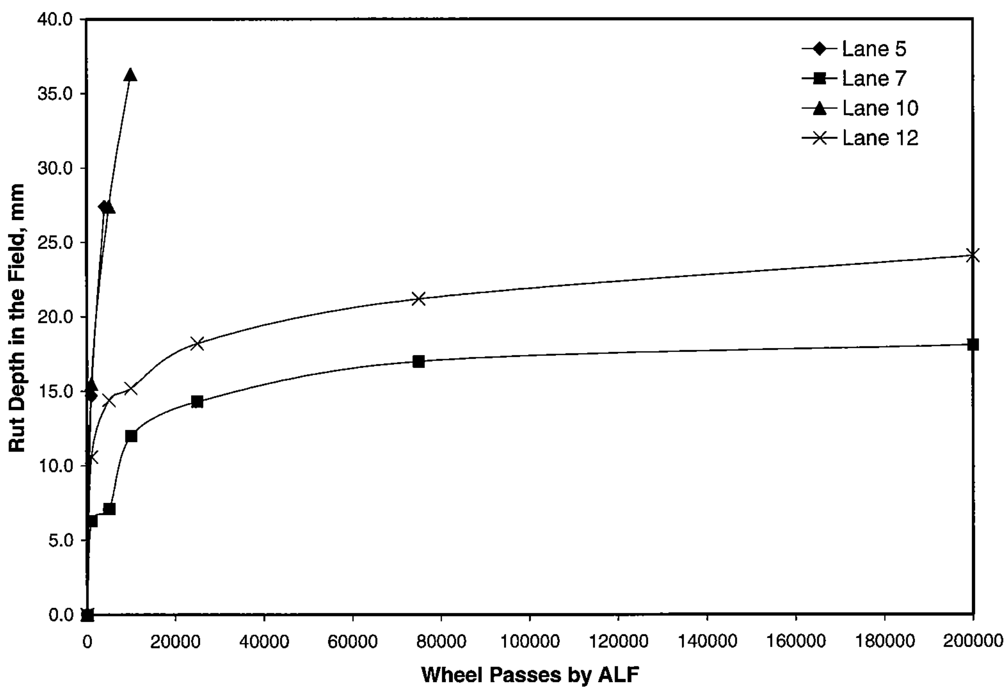


Figure 5. Rutting in the ALF.

TABLE 7 Conversion of ALF rut depths to equivalent rut depths after 20 years and 10,000,000 ESALs (*21*)

Mix	Rut Depth in ALF, mm (No. of ALF Passes)	Rutting Rate, mm/year	Rut Depth after 10,000,000 ESALs, mm
Lane 5	27.4 (4,000)	68.5	1,370
Lane 7	18.1 (200,000)	0.9	18
Lane 10	36.3 (10,000)	26	520
Lane 12	24.1 (200,000)	1.2	24

TABLE 8 Rutting performance of TFHRC ALF mixes (22)

Mixture		AC-10 Lane 5	AC-20 Lane 10	Styrelf 88 Lane 7	AC-20 Lane 12
French LCPC at 60°C					
Cycles		Rut Depth, %			
	300	3.0	2.6	1.8	2.4
	1,000	4.0	3.2	2.2	3.1
	3,000	5.3	4.1	3.0	4.0
	10,000	9.2	4.9	3.2	6.2
	30,000 (spec)	13.8	6.4	3.7	10.9
Georgia LWT at 40°C		Rut Depth, mm			
8,000 Cycles		5.4	3.7	1.9	3.5
Hamburg WTD at 50°C					
Cycles		Rut Depth, mm			
	10,000	22.8	6.8	2.6	4.9
	20,000	>30	8.5	2.8	8.6

3.2 MATERIALS CHARACTERIZATION

For each of the mixtures from the three field projects, a full materials characterization was performed. Tests conducted in this characterization are outlined in Figure 6.

3.2.1 Mixes from WesTrack

Materials obtained from WesTrack included five aggregate stockpiles and one asphalt binder. Hydrated lime was also used in each mix at 1.5 percent by total aggregate mass. Four of the aggregate stockpiles were from the Dayton gravel source and one was a Wadsworth sand. Tables 9 and 10 present properties of the coarse and fine aggregates, respectively. The asphalt binder was a PG 64-22. Results of Superpave binder testing on this source are presented in Table 11. Based upon work conducted in NCHRP Project 9-19, mixing and compaction temperatures for the binder were 152 and 141°C (305 and 282°F), respectively.

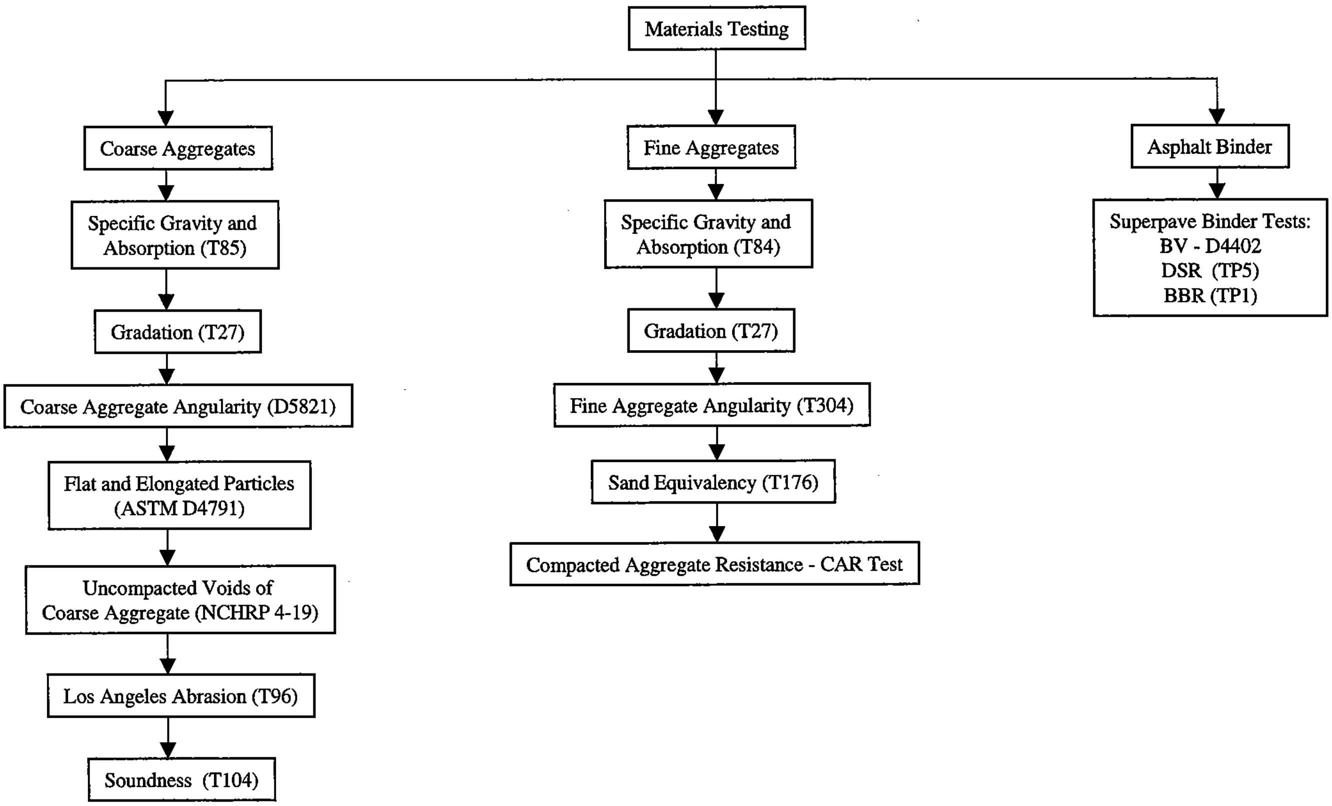
3.2.2 Mixes from MnRoad

Materials obtained from MnRoad included three aggregate stockpiles and two asphalt binders. All three aggregate stockpiles were obtained from the Crow River source. Tables 12 and 13 present properties of coarse and fine aggregates, respectively. The two asphalt binders included a viscosity-graded AC-20 and a 120/150 penetration-graded binder. Tables 14 and 15 present results of the performance grading on these two asphalts. The AC-20 would have a performance grade of PG 64-22; the 120/150 penetration-graded binder would have a performance grade of PG 58-28.

3.2.3 Mixes from FHWA ALF

The No. 68 traprock stockpile used in the original construction of the ALF sections had been depleted at TFHRC prior to the initiation of NCHRP Project 9-17. Unfortunately, the quarry from which the original No. 68 stockpile was obtained had changed crushers. FHWA located a quarry that was near the original quarry (approximately 0.5 miles) and that used a cone crusher similar to that used by the original source. Based on the following observations, this new source was selected for the No. 68 stockpile:

1. The volumetric properties of compacted HMA samples containing the original No. 68 (tested in 1996) using the SGC were reasonably close to those of HMA samples containing the new No. 68 stockpile tested in October 1999.
2. The flat and elongated particle counts of the material retained on the 9.5-mm sieve were similar for both the original and new No. 68 stockpiles.
3. Diabase aggregate is usually reasonably uniform (unlike sedimentary rocks, which can vary from ledge to ledge), and the original and new No. 68 stockpiles were located within 0.5 miles of each other. Also, both stockpiles were crushed by a cone crusher.
4. The four ALF sections selected for NCHRP Project 9-17 will contain the same aggregate and gradation, so the only variable is the asphalt binder type. Therefore, even if one component of the total aggregate is slightly different, the rut depths and ranking of the four mixes will be influenced primarily by the binder type.
5. The NCHRP Project 9-19 researcher also used the new No. 68 stockpile material.



NOTE: Numbers within parentheses refer to AASHTO or ASTM Standard Procedures.

Figure 6. Materials characterization testing.

TABLE 9 Coarse aggregate properties for WesTrack materials

Parameter	Dayton Gravel Source		
	¾ in. Gravel	½ in. Gravel	_ in. Gravel
Bulk Sp. Gr. (T85)	2.539	2.553	2.519
Apparent Sp. Gr. (T85)	2.701	2.679	2.643
Absorption (T85)	2.4	1.8	1.9
Flat and Elongated (ASTM D4791)			
3:1	*	16	*
5:1	*	1.4	*

* Data not available.

TABLE 10 Fine aggregate properties for WesTrack materials

Parameter	Dayton Rock Dust	Wadsworth Sand
Bulk Sp. Gr. (T84)	2.438	2.569
Apparent Sp. Gr. (T84)	2.711	2.727
Absorption (T84)	4.1	2.3
Sand Equivalency (T176)	52	70

Tables 16 and 17 present properties of the coarse and fine aggregates used in the ALF mixes, respectively. All coarse aggregates were a diabase (traprock). The three asphalt binders included an AC-10 and AC-20 and a polymer-modified binder. Tables 18 through 20 present results of the performance grading of the binders. The AC-10, AC-20, and SBS polymer-modified binder have performance grades of PG 58-22, PG 64-22, and PG 82-22, respectively.

TABLE 11 Superpave binder properties for WesTrack materials

Aging	Test Method	Test Temperature	Test Parameter	Value
	Flash Point (T48)	Not applicable	Not applicable	276°C
	Viscosity (D4402)	135°C	Not applicable	0.31 Pa·s
Original	DSR (TP5)	64°C	$G^*/\sin\delta$	1.141 kPa
	DSR (TP5)	64°C	$G^*/\sin\delta$	2.637 kPa
RTFO	Mass Loss (T240)	Not applicable	Not applicable	0.25%
	DSR (TP5)	25°C	$G^*\sin\delta$	4270 kPa
	BBR (TP1)	–12°C	Stiffness	216.8 MPa
RTFO + PAV	BBR (TP1)	–12°C	*m*-value	0.315

NOTES: DSR—dynamic shear rheometer; RTFO—rolling thin film oven; BBR—bending beam rheometer; PAV—pressure aging vessel.

TABLE 12 Coarse aggregate properties for MnRoad materials

Parameter		Crow River Source	
		Crow River Coarses	CA-50
Bulk Sp. Gr. (T85)		2.680	2.712
Apparent Sp. Gr. (T85)		2.760	2.738
Absorption (T85), %		1.2	0.3
Coarse Agg. Angularity 1F (D5821)		61.2	100.0
Uncompacted Voids of C.A. (TP 56-97), %		47.9	48.2
Los Angeles Abrasion (T96), % Loss		27	18
Flat and Elongated (ASTM D4791)			
	3:1	3.0	15
	5:1	5.1	0.7

TABLE 13 Fine aggregate properties for MnRoad materials

Parameter	Crow River Fines
Bulk Sp. Gr. (T84)	2.577
Apparent Sp. Gr. (T84)	2.688
Absorption (T84)	1.6
Fine Agg. Angularity (T304)	41.8
Sand Equivalency (T176)	60

TABLE 14 Superpave binder properties for MnRoad AC-20 asphalt binder

Aging	Test Method	Test Temperature	Test Parameter	Value
	Flash Point (T48)	Not applicable	Not applicable	221°C
	Viscosity (D4402)	135°C	Not applicable	0.41 Pa·s
Original	DSR (TP5)	64°C	$G^*/\sin\delta$	1.055 kPa
	DSR (TP5)	64°C	$G^*/\sin\delta$	2.542 kPa
RTFO	Mass Loss (T240)	Not applicable	Not applicable	0.16%
	DSR (TP5)	25°C	$G^*\sin\delta$	3762 kPa
	BBR (TP1)	–12°C	Stiffness	164 MPa
RTFO + PAV	BBR (TP1)	–12°C	*m*-value	0.333

NOTES: DSR—dynamic shear rheometer; RTFO—rolling thin film oven; BBR—bending beam rheometer; PAV—pressure aging vessel.

TABLE 15 Superpave binder properties for MnRoad 120/150 asphalt binder

Aging	Test Method	Test Temperature	Test Parameter	Value
Original	Flash Point (T48)	Not applicable	Not applicable	290°C
	Viscosity (D4402)	135°C	Not applicable	0.29 Pa·s
	DSR (TP5)	58°C	$G^*/\sin\delta$	2.29 kPa
RTFO	DSR (TP5)	58°C	$G^*/\sin\delta$	5.609 kPa
	Mass Loss (T240)	Not applicable	Not applicable	0.55%
RTFO + PAV	DSR (TP5)	19°C	$G^*\sin\delta$	TBD
	BBR (TP1)	−18°C	stiffness	133 MPa
	BBR (TP1)	−18°C	*m*-value	0.326

NOTES: DSR—dynamic shear rheometer; RTFO—rolling thin film oven; BBR—bending beam rheometer; PAV—pressure aging vessel; TBD—to be determined.

TABLE 16 Coarse aggregate properties for TFHRC ALF materials

Parameter	No. 357	No. 68	No. 8
Bulk Sp. Gr. (T85)	2.962	2.926	2.956
Apparent Sp. Gr. (T85)	3.012	3.011	3.035
Absorption (T85), %	0.6	1.0	0.9
Uncompacted Voids of C.A. (TP 56-97), %	50.8	49.4	Not tested
Los Angeles Abrasion (T96), % Loss	19	15	12
Flat and Elongated 3:1, %	20	15	Not tested
Flat and Elongated 5:1, %	7	1	Not tested

TABLE 17 Fine aggregate properties for TFHRC ALF materials

Parameter	Natural Sand	Traprock No. 10
Bulk Sp. Gr. (T84)	2.578	2.832
Apparent Sp. Gr. (T84)	2.654	2.997
Absorption (T84)	1.10	1.9
Fine Aggregate Angularity (T304)	45.9	47.7
Sand Equivalency (T176)	77	67

TABLE 18 Superpave binder properties for TFHRC ALF AC-10 asphalt binder

Aging	Test Method	Test Temperature	Test Parameter	Value
Original	Flash Point (T48)	Not applicable	Not applicable	280
	Viscosity (D4402)	135°C	Not applicable	312 cP
	DSR (TP5)	58°C	$G^*/\sin\delta$	1.500 kPa
RTFO	DSR (TP5)	58°C	$G^*/\sin\delta$	4.109 kPa
	Mass Loss (T240)	Not applicable	Not applicable	0.35%
RTFO + PAV	DSR (TP5)	22°C	$G^*\sin\delta$	2280 kPa
	BBR (TP1)	−12°C	Stiffness	82 MPa
	BBR (TP1)	−12°C	*m*-value	0.338

NOTES: DSR—dynamic shear rheometer; RTFO—rolling thin film oven; BBR—bending beam rheometer; PAV—pressure aging vessel.

TABLE 19 Superpave binder properties for TFHRC ALF AC-20 asphalt binder

Aging	Test Method	Test Temperature	Test Parameter	Value
	Flash Point (T48)	Not applicable	Not applicable	309
	Viscosity (D4402)	135°C	Not applicable	425 cP
Original	DSR (TP5)	64°C	$G^*/\sin\delta$	1.517 kPa
	DSR (TP5)	64°C	$G^*/\sin\delta$	4.332 kPa
RTFO	Mass Loss (T240)	Not applicable	Not applicable	0.38%
	DSR (TP5)	25°C	$G^*\sin\delta$	2574 kPa
	BBR (TP1)	–12°C	Stiffness	126 MPa
RTFO + PAV	BBR (TP1)	–12°C	m-value	0.330

NOTES: DSR—dynamic shear rheometer; RTFO—rolling thin film oven; BBR—bending beam rheometer; PAV—pressure aging vessel.

TABLE 20 Superpave binder properties for TFHRC ALF polymer-modified asphalt binder

Aging	Test Method	Test Temperature	Test Parameter	Value
	Flash Point (T48)	Not applicable	Not applicable	Not applicable
	Viscosity (D4402)	135°C	Not applicable	2108 cP
Original	DSR (TP5)	82°C	$G^*/\sin\delta$	1.362 kPa
	DSR (TP5)	82°C	$G^*/\sin\delta$	3.107 kPa
RTFO	Mass Loss (T240)	Not applicable	Not applicable	Not applicable
	DSR (TP5)	34°C	$G^*\sin\delta$	1130 kPa
	BBR (TP1)	–12°C	Stiffness	100 MPa
RTFO + PAV	BBR (TP1)	–12°C	m-value	0.302

NOTES: DSR—dynamic shear rheometer; RTFO—rolling thin film oven; BBR—bending beam rheometer; PAV—pressure aging vessel.

CHAPTER 4

EXPERIMENTAL PLAN (PHASE I)

4.1 INTRODUCTION

Based upon the review of literature and guidance from the project panel, a controlled laboratory experiment was designed using the materials described in Chapter 3. The primary objectives of the experiment were to evaluate variables that could potentially influence the ability of the APA to predict the rutting potential of HMA mixtures in the field and to select the combination of variables that best predicts rutting potential.

The overall research approach is shown in Figure 7. After completion of the main experiment, the data were analyzed and conclusions drawn about the ability of the APA to predict field rut depths. The following sections describe the full-factorial main experiment.

4.2 EXPERIMENTAL PLAN USING FIELD MIXES OF KNOWN PERFORMANCE

Four factors were included along with the 10 mixes described in Chapter 3 within the plan of the experiment. These factors and their levels are as follows:

- **Specimen Type:** Beams compacted with an AVC; cylinders compacted with an SGC.
- **Hose Diameter:** The standard hose diameter of 25 mm (outside diameter); hose with a diameter of 38 mm (outside diameter).
- **Test Temperature:** High temperature of standard performance grade based on climate; 6°C higher than high temperature of standard performance grade.
- **Air Void Content:** 4.0 ± 0.5 percent; 7.0 ± 0.5 percent.

Table 21 gives the test matrix for the experimental plan for Task 4. A wheel load and hose pressure of 534 N (120 lb) and 827 kPa (120 psi), respectively, were used for the entire study. Test temperatures used for each of the 10 mixes are presented in Table 22.

The project panel permitted air void contents slightly higher than 4 percent for beam samples due to the difficulty in compacting beam samples to 4-percent air voids in the AVC. Based upon work with the 10 mixes selected for NCHRP Project 9-17, a target air void content of 5.0 ± 0.5 was selected for beam samples.

This experiment involved 160 factor-level combinations (2 sample types × 2 hose diameters × 2 test temperatures × 2 air void contents × 10 mixes). Three replicates of each factor-level combination were tested. A single replicate was considered the average rut depth for two cylinders or one beam. Testing was conducted on mixes fabricated from original materials, proportioned to meet in-place properties, mixed in the laboratory, and subjected to short-term aging per AASHTO TP 2-96.

4.2.1 Statistical Analysis

A statistical approach to evaluating the effectiveness of the APA for predicting rut depths in the field is very difficult. Numerous combinations of laboratory testing conditions were utilized to compare field and laboratory rut depths. Previous research studies that have evaluated the ability of laboratory LWTs to predict field rutting have shown that the exact magnitude of rutting in the field may not be accurately predicted by laboratory testing. However, some studies have shown that strong correlations exist between laboratory and field rut depths.

A statistical approach that can provide a definitive answer as to which combination of laboratory testing conditions (factor-level combinations) provides the best correlation is not available. The type of data generated from a study of this nature essentially compares one field data point to one laboratory data point, both of which are independent. So for a given field project, a single mean laboratory rut depth can only be compared with a single mean field rut depth.

Therefore, the primary analysis tool selected for comparing laboratory and field rut depths was a simple correlation/regression analysis. For each factor-level combination investigated in the APA, a scatter plot was developed that described the results of laboratory and field rut depths. Each plot reflected actual field rutting versus laboratory rut depth for a given factor-level combination for a given pavement. A correlation/regression analysis was then conducted on the data in order to determine the best fit line and the coefficient of correlation (R^2).

Selection of the optimum factor-level combination for testing conditions in the APA was based upon the highest R^2

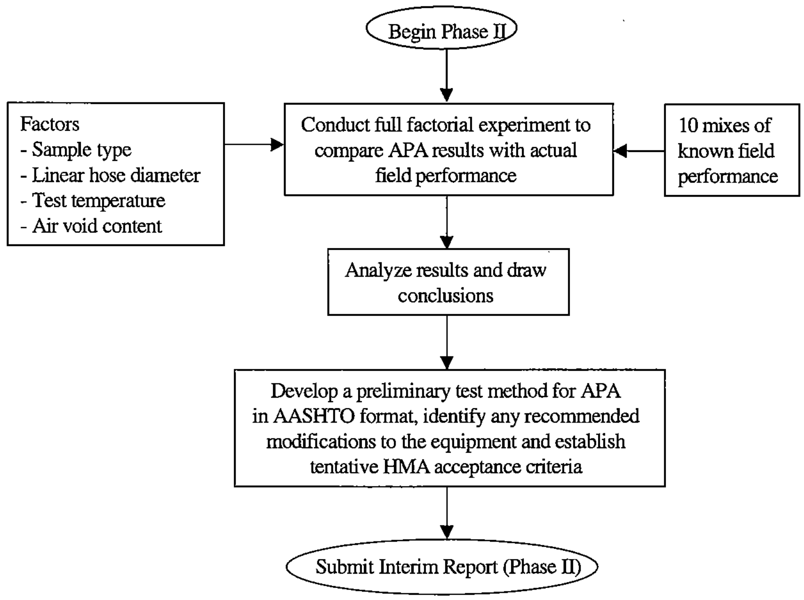


Figure 7. Overall research approach (all materials used are discussed in Chapter 3).

value obtained from the regression analyses. If one combination consistently shows a higher R^2 value than did all other combinations, it would be selected and included in the tentative standard procedure. However, the potential existed that several factor-level combinations would have approximately the same R^2 value. Therefore, several additional statistical analyses were conducted in order to break any ties for the optimum factor-level combination.

First, a Duncan's Multiple Range Test (DMRT) was conducted for the data from each project (i.e., WesTrack, MnRoad, and FHWA ALF). This methodology was selected because the climatic and traffic conditions for pavements within a given project should be identical (hence, the reasoning behind the selection of multiple pavements from each project). The only differences between the test sections from a given project should be mix type and degree of rutting. Therefore, a DMRT ranking of laboratory data should match the field data. If this does not occur for a given factor-level combination, it could be disregarded. For instance, assume the three field pavements from MnRoad have a high, low, and medium degree of rutting. A DMRT ranking of the laboratory data should replicate the field ranking.

Second, individual correlation/regression analyses were conducted for data from each project (i.e., WesTrack, MnRoad, and FHWA ALF). This methodology was selected because the climatic and traffic conditions should be identical within a given project. Therefore, these individual correlation/regression analyses for each project may identify traffic or climatic characteristics that influence the ability of an APA to predict pavement performance in terms of rutting.

Third, once a number of factor-level combinations were discounted due to improperly ranking field mixes or because of low R^2 values, the variability of the remaining test procedures was evaluated. The first procedure for this analysis was to determine the coefficient of variation within selected test combinations. Three replicates represent a single test within the APA. The mean and standard deviation were calculated for each APA test using these replicates. The coefficient of variation (COV) is a measure of variability and is defined as the standard deviation divided by the mean and expressed as a percentage. As the COV increases, more variability is expected in a test procedure. A typical acceptable value for COV is 10 percent for tests of HMA. The second method of evaluating variability was to conduct replicate tests on selected combinations. Replicate tests provided an evaluation of the repeatability of a test procedure.

4.2.2 Evaluation of Boundary Conditions on APA Test Specimens

This limited evaluation had the goal of investigating the loading conditions in the APA to demonstrate the degree of interaction present in the sample from the edge conditions of the sample holding system of plastic or steel restraints. This analysis was conducted as part of the project.

The APA beam was modeled as a solid contained within a stationary boundary which allowed no displacement of the sample in the horizontal direction. A modulus of 100,000 psi

TABLE 21 Experimental design for Task 4 of Phase II

Air Void Contents	Test Temperature	Hose Diameter	Specimen Type	Field Mixes									
				1	2	3	4	5	6	7	8	9	10
			Beam	X	X	X	X	X	X	X	X	X	X
		25 mm	Cylinder	X	X	X	X	X	X	X	X	X	X
			Beam	X	X	X	X	X	X	X	X	X	X
	PG High Temp.	38 mm	Cylinder	X	X	X	X	X	X	X	X	X	X
			Beam	X	X	X	X	X	X	X	X	X	X
		25 mm	Cylinder	X	X	X	X	X	X	X	X	X	X
			Beam	X	X	X	X	X	X	X	X	X	X
7 ± 0.5 percent	PG High Temp. plus 6°C	38 mm	Cylinder	X	X	X	X	X	X	X	X	X	X
			Beam	X	X	X	X	X	X	X	X	X	X
		25 mm	Cylinder	X	X	X	X	X	X	X	X	X	X
			Beam	X	X	X	X	X	X	X	X	X	X
	PG High Temp.	38 mm	Cylinder	X	X	X	X	X	X	X	X	X	X
			Beam	X	X	X	X	X	X	X	X	X	X
		25 mm	Cylinder	X	X	X	X	X	X	X	X	X	X
			Beam	X	X	X	X	X	X	X	X	X	X
4 ± 0.5 percent	PG High Temp. plus 6°C	38 mm	Cylinder	X	X	X	X	X	X	X	X	X	X

NOTE: X—factor-level combinations to be tested.

(689.5 MPa) was selected to model a high temperature of 50°C and to maintain a slightly elevated stiffness level to accentuate edge effects on the spread of load from under the loaded hose. A softer sample would dissipate the load stresses more quickly. The load was applied over a strip 25 mm long and 19.16 mm wide, giving a contact area of 7.29 cm^2. This produces a contact stress of 730 kPa under the hose.

The computer program ABAQUS was used to model the stress distribution in the typical APA beam. The results indicate that there was very little interaction with the stress and edge of the beams in the vicinity of the wheel load for the given input and boundary conditions. The beam was in a uniform state of stress ahead and behind the wheel, and this same stress level exists for approximately 1 in. adjacent to the edge of the sample. Thus, the moving stress pulse is uniformly surrounded by a state of stress that does not interact with the sidewalls at all. The depth of the beam could have an impact because the stress bulb from the hose interacts with the bottom of the beam, and varying support along the bottom of the beam could produce a different stress pattern, although the stress level here is only approximately $^1/_{10}$ of the applied contact stress. This stress level would be insignificant in producing variability in the observed rutting on the sample surface.

TABLE 22 Test temperatures for 10 field mixes

Project	Mix	Test Temp, °C
	Section 15	64 and 70
	Section 19	64 and 70
WesTrack	Section 24	64 and 70
	Cell 16	58 and 64
	Cell 20	58 and 64
MnRoad	Cell 21	58 and 64
	Lane 5	58 and 64
	Lane 10	58 and 64
	Lane 7	58 and 64
TFHRC-ALF	Lane 12	58 and 64

It is highly unlikely that edge effects play a part in any variability seen in APA results. If higher loads and contact stresses were used, the bottom effects might play a role in altering the development of rutting and in compounding any direct comparison of increased loads and their impact on rut development in the APA unless special precautions are taken to ensure consistent bonding of the samples and the bottom of the mold. The edge of the samples will most likely not influence results even when higher contact stresses are used.

This distribution of the stresses indicates that gyratory samples will experience end effects until the load has traveled 25 mm onto the sample; thus, rut depth measurements should be taken only over the center 75 mm of the gyratory sample. It would not be expected that significant differences in rut measurements would be attributable to the edge effects for the two different sample types, and any differences are more likely to be the result of the different compaction methods producing different aggregate structures in the mixture.

4.3 ADDITIONAL TESTING OF 10 FIELD MIXES

Of the 10 field mixes, 7 were tested using the Hamburg Wheel Tracking Device (HWTD) owned by APAC Materials Services in Smyrna, Georgia. Sufficient materials were not available to test the three WesTrack mixes. The HWTD was selected for this comparison testing because of its convenience and location.

Field mixes were also tested with the simple performance test identified by NCHRP Project 9-19. NCHRP provided the researchers with the test results, and a comparison among the three simple performance tests and the APA was conducted.

CHAPTER 5

TEST RESULTS AND ANALYSIS (PHASE II)

5.1 INTRODUCTION

This chapter presents the test results and analysis of the full factorial laboratory experiment conducted in Phase II. This chapter is composed of three main sections. The first section discusses test results and analyses for each of the full-scale pavement studies individually. The second section discusses test results and analyses conducted using all three pavement studies collectively. The final section discusses the set of APA testing conditions selected for validation in Phase III, justification for the conditions, and tentative acceptance criteria based upon the selected test procedure(s).

Sixteen different laboratory testing combinations will be discussed. Each combination includes an air void content, test temperature, hose diameter, and sample type. For simplicity's sake, a nomenclature for each testing combination was developed (Figure 8). As shown in the figure, the first character represents the air void content of the sample. Air void contents were 7, 5, and 4 percent, depending upon the level and sample type. The second and third characters represent test temperature (either PG or PG+6°C). Hose diameter is depicted by the fourth character: an "L" for the large diameter hose or an "S" for the standard hose. The final character represents the sample type: a "B" represents a beam sample, and a "C" represents a cylindrical sample.

5.2 TEST RESULTS AND ANALYSIS FOR INDIVIDUAL PAVEMENT STUDIES

This section discusses test results and analysis for each of the three full-scale pavement studies individually.

5.2.1 Mixes from WesTrack

As stated previously, three mixes were evaluated in this study for the WesTrack field experiment. Test Sections 15 (fine gradation), 19 (fine-plus gradation), and 24 (coarse gradation) were selected. All three of these sections were placed at "optimum" asphalt content and in-place air voids. Results of APA testing on the WesTrack sections are presented in Table 23. Rut depths in this table represent manual measurements after 10,000 cycles in the APA.

Based upon actual field rutting (illustrated in Figure 3), Section 15 had the lowest amount of rutting in the field, followed by Section 19 and Section 24, respectively. Figure 3 suggests that the difference in the level of rutting was significant among the three sections.

DMRTs were conducted for each factor-level combination (16 APA testing conditions) in order to rank the laboratory results for the three section mixes (level of significance [α] = 0.05). All three replicates within a test were used for this analysis. Table 24 presents the results of the DMRT rankings for the laboratory mixes. Within this table, the field mixes were also assigned rankings. Since replicate field rut depths for each section were not available, a statistical ranking was not performed. The field rankings shown in Table 24 are based upon Figure 3. Because Section 24 failed and was rehabilitated after 2.8 million ESALs, rut depths at this traffic level were determined for the three sections from the data in Figure 3. Based upon the rut depths at 2.8 million ESALs, Section 24 obviously had a significantly higher amount of rutting than Sections 15 and 19. Field rutting in Sections 15 and 19 may not be statistically different, but are most likely different.

As shown in Table 24, only one laboratory testing combination had similar statistical rankings as the field rut depths: 5PGSB. Obviously from Table 24, the 7-percent combinations did not properly rank the mixes because Section 24 did not have the highest laboratory rut depths. All but two combinations (4P+LC and 5P+LB) of the lower air void content samples correctly showed that Section 24 had the highest laboratory rut depths. Three other laboratory testing combinations showed a similar trend as the field performance: 4PGSC, 5PGLB, and 5P+SB. A similar trend infers that Section 24 showed the highest laboratory rut depth and that Section 15 had the lowest.

In order to determine the combination of laboratory testing conditions that best predicted field rutting, a regression of field rutting versus laboratory rut depth was developed for each combination. Field rut depths at 2.8 million ESALs were again used in this analysis. In order to accurately compare the different laboratory combinations with the WesTrack field rutting, a similar traffic level was needed.

Table 25 presents R^2 values for each laboratory testing combination versus field rut depth. Within this table, some R^2

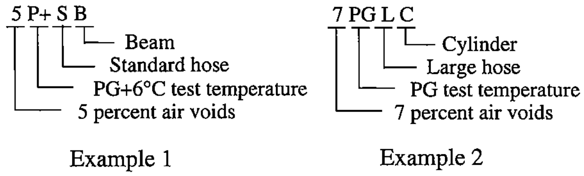


Figure 8. Nomenclature to describe different laboratory testing conditions.

values show a negative value. These laboratory combinations had a relationship with field rutting that was opposite of what would be considered acceptable. In other words, laboratory rut depths decreased as field rutting increased. Interestingly, all but one of the 7-percent air void samples had negative slopes.

The information in Table 25 was used to determine the six testing combinations with the highest R^2 value. These combinations are shown in Table 26. All six combinations shown in this table had the low level for the air void content factor (4 percent for cylinders and 5 percent for beams). Four of the six combinations used the standard performance-grade high temperature, and four also used the standard hose. Half of the top six combinations used a beam sample, and half used a cylinder.

The results shown in Table 26 are interesting in that the laboratory combination that correctly ranked field performance did not have the highest R^2 value. Also, the 5PGLB combination had an R^2 value of 1.000 even though this combination did not statistically rank the field mixes correctly. Two of the six combinations did not rank the mixes correctly or show the same trend as field rutting: 4P+SC and 4PGLC. Laboratory rut depths for the three mixes when using either of these combinations varied less than 2 mm.

In Tables 24 and 26, it appears that four laboratory testing combinations can be considered from the WesTrack data: 5PGLB, 5P+SB, 4PGSC, and 5PGSB. The relationship between field and laboratory rut depths for these four combinations are illustrated in Figure 9. All four of the relationships shown in this figure are acceptable because all have R^2 values above 0.83 and the slope of the lines is what would be expected. However, the 5PGLB and 4PGSC combinations have similar steep slopes. Slopes that are too steep are detrimental to the overall objective of this study because little difference in laboratory rut depths would be expected between good and poor performing mixes.

One of the purposes of selecting Sections 15, 19, and 24 from the WesTrack experiment was that the selection allowed for a comparison of different Superpave gradation types. All three sections used the same asphalt binder (although the asphalt content varied). Figure 10 illustrates the effect of gradation type on laboratory rutting for the 5PGSB and 5P+SB combinations. This figure shows that the fine gradation (Section 15) had the lowest laboratory rut depth. The coarse gradation (Section 24) had the highest laboratory rut depths. This trend in laboratory results is similar to that shown in the field; therefore, these two test combinations differentiated between gradations.

TABLE 23 Average rut depths for WesTrack sections (manual reading)

Air Voids	Test Temp, °C	Hose Diameter	Specimen Type	Rut Depth @ 10,000 cycles, mm		
				Section 15	Section 19	Section 24
			Cylinder	10.13	6.12	8.17
		Standard	Beam	8.52	9.14	8.70
			Cylinder	8.30	4.19	4.05
	64 (PG)	Larger	Beam	6.46	7.99	6.39
			Cylinder	12.78	8.61	10.56
		Standard	Beam	12.65	11.14	11.69
			Cylinder	12.01	7.22	4.57
7%	70 (PG+6)	Larger	Beam	9.33	6.21	8.40
			Cylinder	6.64	7.76	8.27
		Standard	Beam	6.08	10.84	13.33
			Cylinder	6.36	5.35	7.54
	64 (PG)	Larger	Beam	5.10	5.76	7.08
			Cylinder	8.81	8.80	9.29
		Standard	Beam	11.57	13.07	14.88
			Cylinder	7.83	6.10	5.47
4%*	70 (PG+6)	Larger	Beam	5.64	6.87	5.24

*Beam samples compacted to 5.0 ± 0.5% air voids.

TABLE 24 Comparison of field and laboratory rut depth rankings for WesTrack

Air Voids	Test Temp., °C	Hose Diameter	Specimen Type	Field Rank* (Section No.)	Lab Rank* (Section No.)
7%	64 (PG)	Standard	Cylinder	24, 19, 15	15, 24, 19
			Beam	24, 19, 15	*19, 24, 15*
		Larger	Cylinder	24, 19, 15	15, *19, 24*
			Beam	24, 19, 15	*19, 15, 24*
	70 (PG+6)	Standard	Cylinder	24, 19, 15	15, 24, 19
			Beam	24, 19, 15	*15, 24, 19*
		Larger	Cylinder	24, 19, 15	15, 19, 24
			Beam	24, 19, 15	*15,* ***24,19***
4% [c]	64 (PG)	Standard	Cylinder	24, 19, 15	*24, 19, 15* [a]
			Beam	24, 19, 15	24, 19, 15 [b]
		Larger	Cylinder	24, 19, 15	*24, 15, 19*
			Beam	24, 19, 15	*24,* ***19, 15*** [a]
	70 (PG+6)	Standard	Cylinder	24, 19, 15	*24, 15, 19*
			Beam	24, 19, 15	24, *19, 15* [a]
		Larger	Cylinder	24, 19, 15	*15,* ***19, 24***
			Beam	24, 19, 15	*19, 15, 24*

* Sections in order of decreasing field rut depth. Italics and bold indicate statistically similar rankings. Lab rankings based upon Duncan's Multiple Range Test.
[a] Laboratory results show the same trend as field rutting.
[b] Laboratory rankings similar to field rankings.
[c] Beam samples compacted to 5.0 ± 0.5%.

TABLE 25 R^2 values for WesTrack factor-level combinations

Air Voids	Test Temp, °C	Hose Diameter	Specimen Type	R^2 *
7%	60 (PG)	Standard	Cylinder	–0.081
			Beam	0.020
		Larger	Cylinder	–0.650
			Beam	–0.010
	70 (PG + 6)	Standard	Cylinder	–0.119
			Beam	–0.239
		Larger	Cylinder	–0.967
			Beam	–0.037
4% [c]	64 (PG)	Standard	Cylinder	0.856 [a]
			Beam	0.835 [b]
		Larger	Cylinder	0.386
			Beam	1.000 [a]
	70 (PG + 6)	Standard	Cylinder	0.855
			Beam	0.982 [b]
		Larger	Cylinder	–0.866
			Beam	–0.164

*A negative sign indicates that the relationship between field and laboratory rutting was opposite of what would be considered acceptable.
[a] Laboratory results show the same trend as field rutting.
[b] Laboratory rankings similar to field rankings.
[c] Beam samples compacted to 5.0±0.5% air voids.

TABLE 26 Six highest R^2 values for WesTrack

Air Voids	Test Temp	Hose Diameter	Specimen Type	R^2
5%	PG	Large	Beam	1.000 [a]
5%	PG+6	Standard	Beam	0.982 [a]
4%	PG	Standard	Cylinder	0.856 [a]
4%	PG+6	Standard	Cylinder	0.855
5%	PG	Standard	Beam	0.835 [b]
4%	PG	Large	Cylinder	0.386

[a] Laboratory results show the same trend as field rutting.
[b] Laboratory results ranked statistically similar to field rutting.

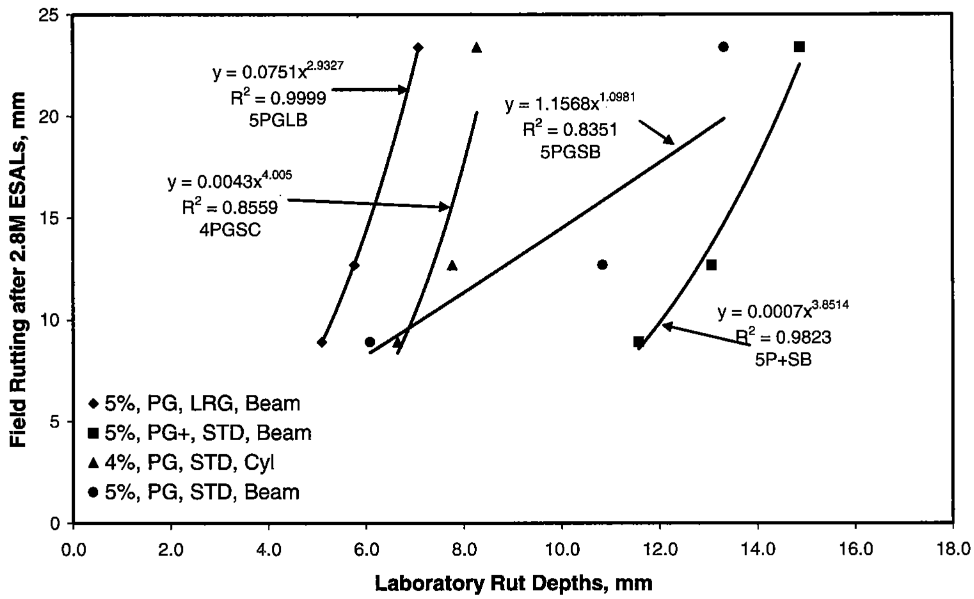


Figure 9. Four potential lab testing combinations for WesTrack.

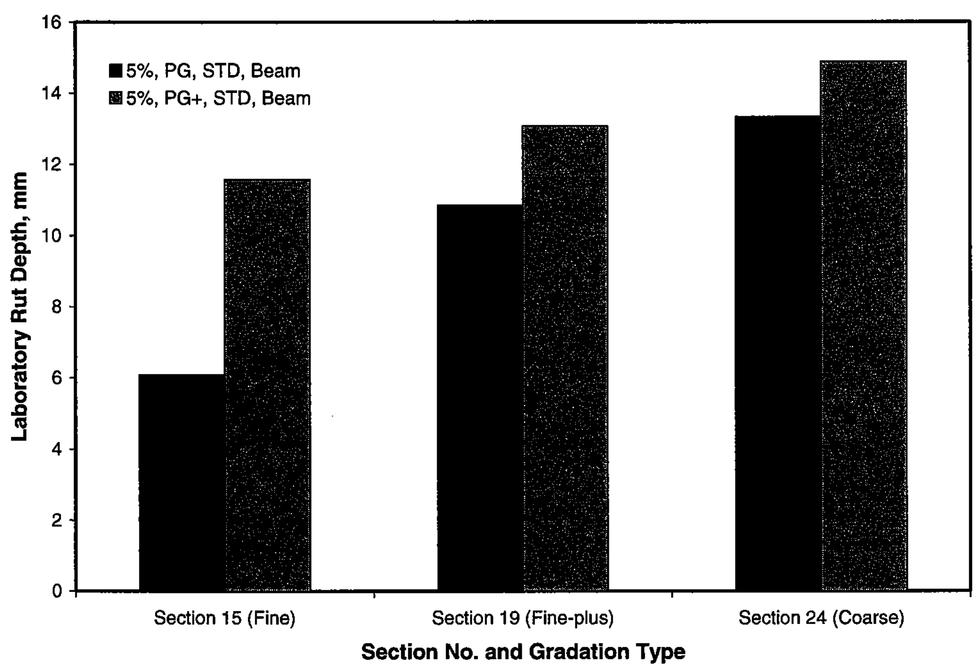


Figure 10. Comparison of WesTrack gradations.

5.2.2 Mixes from MnRoad

Three mixes were included from the MnRoad field experiment: Cell 16, Cell 20, and Cell 21. All three cells had identical gradations, but differed by asphalt binder type, asphalt content, or both. Results of APA testing for these cells are presented in Table 27. Rut depths in this table represent manual rut depth measurements after 10,000 cycles in the APA.

Based upon actual field rutting (Figure 4), Cell 16 had the lowest amount of rutting, followed by Cell 21 and Cell 20, respectively. Replicate field rut depth measurements for each cell were available (10 per cell). The replicate data were used to rank (by DMRT) the field performance for each cell after 2 million ESALs (the point at which Cell 20 began tertiary flow and was rehabilitated). Results of this analysis showed that each of the cells had significantly different field rut depths.

DMRT rankings of laboratory rut depths were also conducted for each of the testing combinations. Results of the rankings are presented in Table 28 along with the field performance rankings. Only one laboratory combination (5PGLB) correctly ranked the mixes as to field performance. Six laboratory combinations did show a similar trend as field rutting in that Cell 20 had the highest magnitude of rutting and Cell 16 had the lowest: 7PGLC, 7PGLB, 4PGSC, 4PGLC, 4P+SC, and 5P+LB. Of the seven laboratory testing combinations that ranked properly or showed the same trend as field rutting, five had the low level of air void content (4 percent for cylinders and 5 percent for beams). Five also used the high temperature of the performance grade (58°C). Four combinations had cylindrical samples.

The data within Table 27 along with the field rutting in each cell were used to develop relationships between field and laboratory rut depths. Table 29 presents the R^2 values for each factor-level combination when field rutting versus laboratory rut depths are compared. Again, field rut depths are measurements at 2 million ESALs. Collectively, the R^2 values shown in Table 29 are better than those presented for the WesTrack data (Table 25). Only one combination showed a negative slope (5P+SB); however, this combination was one of the two more viable combinations for the WesTrack data.

The seven laboratory testing combinations that had statistically similar rankings or showed a similar trend as field rutting also had the seven highest R^2 values. Table 30 presents the six combinations for the MnRoad mixes with the highest R^2 values. All six had high R^2 values of higher than 8.4. Four of these six combinations used the low level of the air void content factor. Five had a test temperature corresponding to the standard performance-grade high temperature (58°C). Five of the six also were tested with the larger-diameter hose. Similar to the WesTrack data, the top six combinations were evenly distributed between beam and cylindrical sample types.

Figures 11 and 12 illustrate the relationship between field rutting (at 2 million ESALs) and laboratory rut depths for the six combinations shown in Table 30. All six combinations appear to differentiate between the well-performing (Cell 16) and poorly performing (Cell 20) mixes. The 4PGSC and

TABLE 27 Average rut depths for MnRoad sections (manual reading)

Air Voids	Test Temp, °C	Hose Diameter	Specimen Type	Rut Depth @ 10,000 cycles, mm		
				Cell 16	Cell 20	Cell 21
			Cylinder	9.91	17.63	8.08
		Standard	Beam	11.20	12.79 [a]	14.87 [b]
			Cylinder	6.72	12.19	11.52
	58 (PG)	Larger	Beam	8.55	13.58	13.12
			Cylinder	11.18 [c]	21.42	22.37
		Standard	Beam	19.86	24.54	24.55
			Cylinder	7.69	13.47	14.67
7%	64 (PG+6)	Larger	Beam	12.78	18.33	18.68
			Cylinder	5.43	17.03	15.96
		Standard	Beam	7.77	19.89	20.61
			Cylinder	6.72	10.43	8.33
	58 (PG)	Larger	Beam	5.75	11.36	8.74
			Cylinder	14.49	19.44	19.33
		Standard	Beam	29.45	25.54	25.14
			Cylinder	8.52	9.58	10.92
4% [d]	64 (PG+6)	Larger	Beam	8.97	16.09	15.62

[a] Test was terminated after 5,454 cycles because the wheels began riding on the samples.
[b] Test was terminated after 5,722 cycles because the wheels began riding on the samples.
[c] Test was terminated after 5,410 cycles because the wheels began riding on the samples.
[d] Beam samples compacted to 5.0 ± 0.5% air voids.

TABLE 28 Comparison of field and laboratory rut depth rankings for MnRoad

Air Voids	Test Temp., °C	Hose Diameter	Specimen Type	Field Rank*	Lab Rank*
7%	58 (PG)	Standard	Cylinder	20, 21,16	20, *16, 21*
			Beam	20, 21,16	*21, 20, 16*
		Larger	Cylinder	20, 21,16	*20, 21,* 16 [a]
			Beam	20, 21,16	*20, 21,* 16 [a]
	64 (PG+6)	Standard	Cylinder	20, 21,16	*21, 20*, 16
			Beam	20, 21,16	*21, 20*, 16
		Larger	Cylinder	20, 21,16	*21, 20*, 16
			Beam	20, 21,16	*21, 20*, 16
4% [c]	58 (PG)	Standard	Cylinder	20, 21,16	*20, 21,* 16 [a]
			Beam	20, 21,16	*21, 20,* 16
		Larger	Cylinder	20, 21,16	*20, 21,* 16 [a]
			Beam	20, 21,16	*20, 21,* 16 [b]
	64 (PG+6)	Standard	Cylinder	20, 21,16	*20, 21,* 16 [a]
			Beam	20, 21,16	*16, 20, 21*
		Larger	Cylinder	20, 21,16	*21, 20, 16*
			Beam	20, 21,16	*20,* ***21, 16*** [a]

*Cells in order of decreasing field rut depth. Italics and bold indicate statistically similar rankings. Lab and field analysis based upon Duncan's Multiple Range Test.
[a] Laboratory results show similar trend as field rutting.
[b] Laboratory rankings similar to field rankings.
[c] Beam samples compacted to 5.0 ± 0.5%.

TABLE 29 R^2 values for MnRoad factor-level combinations

Air Voids	Test Temp, °C	Hose Diameter	Specimen Type	R^2 *
7%	58 (PG)	Standard	Cylinder	0.429
			Beam	0.287
		Larger	Cylinder	0.876 [a]
			Beam	0.863 [a]
	64 (PG + 6)	Standard	Cylinder	0.769
			Beam	0.813
		Larger	Cylinder	0.711
			Beam	0.779
4% [c]	58 (PG)	Standard	Cylinder	0.852 [a]
			Beam	0.788
		Larger	Cylinder	0.992 [a]
			Beam	0.997 [b]
	64 (PG + 6)	Standard	Cylinder	0.827 [a]
			Beam	–0.739
		Larger	Cylinder	0.291
			Beam	0.848 [a]

* A negative sign indicates that the relationship between field and laboratory rutting was opposite of what would be considered acceptable.
[a] Laboratory results show similar trend as field rutting.
[b] Laboratory rankings similar to field rankings.
[c] Beam samples compacted to 5.0 ± 0.5% air voids.

TABLE 30 Six highest R^2 values for MnRoad

Air Voids	Test Temp	Hose Diameter	Specimen Type	R^2
5%	PG	Large	Beam	0.997 [b]
4%	PG	Large	Cylinder	0.992 [a]
7%	PG	Large	Cylinder	0.876 [a]
7%	PG	Large	Beam	0.863 [a]
4%	PG	Standard	Cylinder	0.852 [a]
5%	PG+6	Large	Beam	0.848 [a]

[a] Laboratory results show similar trend as field rutting.
[b] Laboratory rankings similar to field rankings.

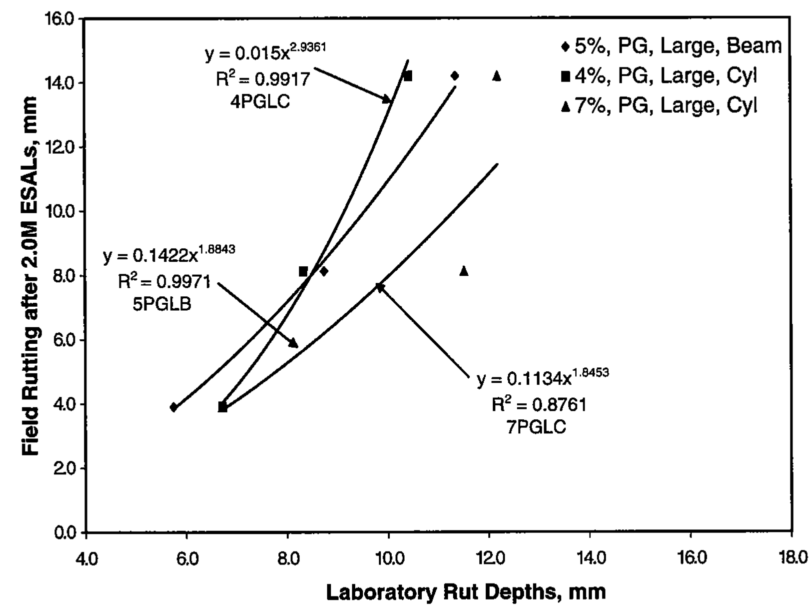


Figure 11. Three highest R² *values for MnRoad mixes.*

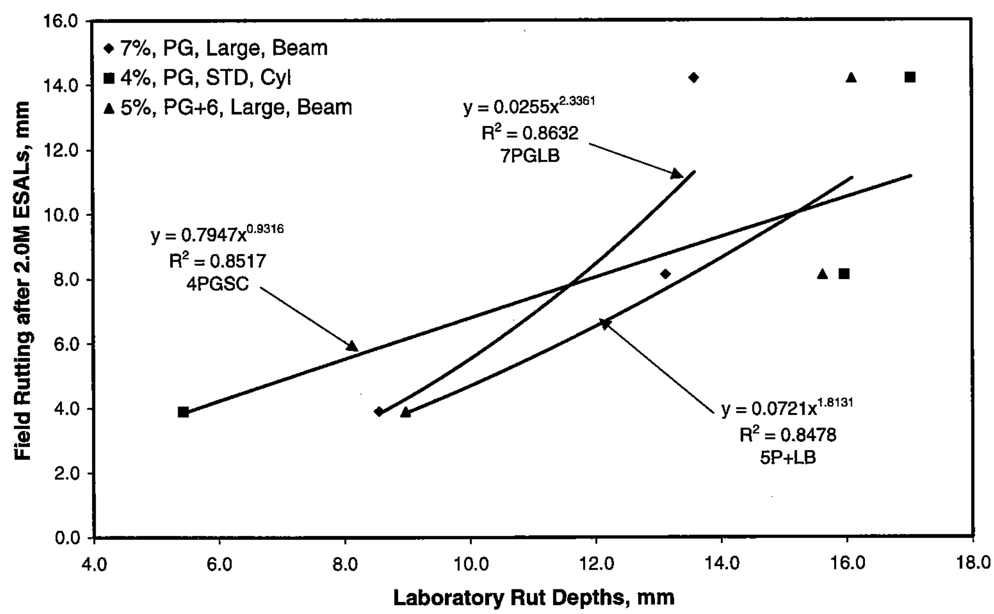


Figure 12. Next three highest R² *values for MnRoad mixes.*

5ΠI+LB combinations do, however, have a flatter slope. The 4PGLC combination has the steepest slope (although not as steep as the 5PGLB and 4PGSC test combinations shown in Figure 9 for the WesTrack data). In the data in Table 30 and Figures 11 and 12, all six combinations appear to be viable options for a recommended test procedure.

The cells selected from MnRoad allow for the effect of asphalt binder type and binder content to be evaluated. As stated in Chapter 3, Cell 16 used an AC-20 asphalt binder while Cells 20 and 21 both used a 120/150 penetration graded binder. Table 3 showed that Cell 16 had an asphalt content of 5.1 percent; Cells 20 and 21 had asphalt contents of 6.1 and 5.9 percent, respectively. Figure 13 illustrates the laboratory rut depths for the three cells and also shows the asphalt binder type and asphalt content. Only the 5PGLB and 4PGLC combinations are shown on this figure. The first observation that stands out about Figure 13 is the difference in rut depths between Cells 20 and 21. Both of these cells used the same type of asphalt binder and only varied by asphalt content (0.2 percent). For both testing conditions shown in the figure, the 0.2-percent difference in asphalt content resulted in approximately a 2-mm difference in rut depth. Cell 16, which had the lowest asphalt content, had the lowest rut depth of the three cells, which is expected.

5.2.3 Mixes from ALF

Four mixes were selected from the ALF experiment conducted at TFHRC. Three of these mixes had identical gradations, but varied by asphalt binder type (Lane 5, Lane 7, and Lane 10). These three mixes were all 19.0-mm NMAS gradations. Lane 12 was the only mix with a different NMAS (37.5 mm). Lane 7 was unique from the other three mixes investigated in this study in that it used a polymer-modified asphalt binder (PG 82-22).

Loading of the experimental pavements at the ALF was characterized by the number of passes of the ALF wheel assembly. As stated in Chapter 3, no conversions from ALF passes to ESALs have been performed; therefore, comparisons between field and laboratory rutting in this section will be for equivalent ALF passes. Stuart et al. (*22*) ranked the ALF surface mixes (Lanes 5, 7, and 10) with respect to rutting in the field. Based upon their rankings, Lanes 5 and 10 had similar magnitudes of rutting and Lane 7 had significantly less rutting, which is similar to that shown in Figure 5. However, the authors did not rank Lane 12 (base mix) with the surface mixes. From Figure 5, it can be surmised that Lane 12 had significantly less rutting than Lanes 5 and 10, but may not be statistically different than Lane 7.

Results of APA testing on the ALF mixes are presented in Table 31. These results represent average, manually determined rut depths after 10,000 cycles in the APA. DMRT rankings were conducted for each testing condition to rank the laboratory rut depths of each lane. Results are presented in Table 32. Rankings of field rut depths are based upon Stuart et al. (*22*) and visual observation of Figure 5. Interestingly, none of the laboratory testing combinations correctly ranked the mixes as to their field rutting. However, six laboratory combinations did show a similar trend as field rutting: 7PGLC, 7P+SB, 4PGLC, 5PGLB, 4P+SC, and 4P+LC. Similar to the WesTrack and MnRoad data, the majority of these

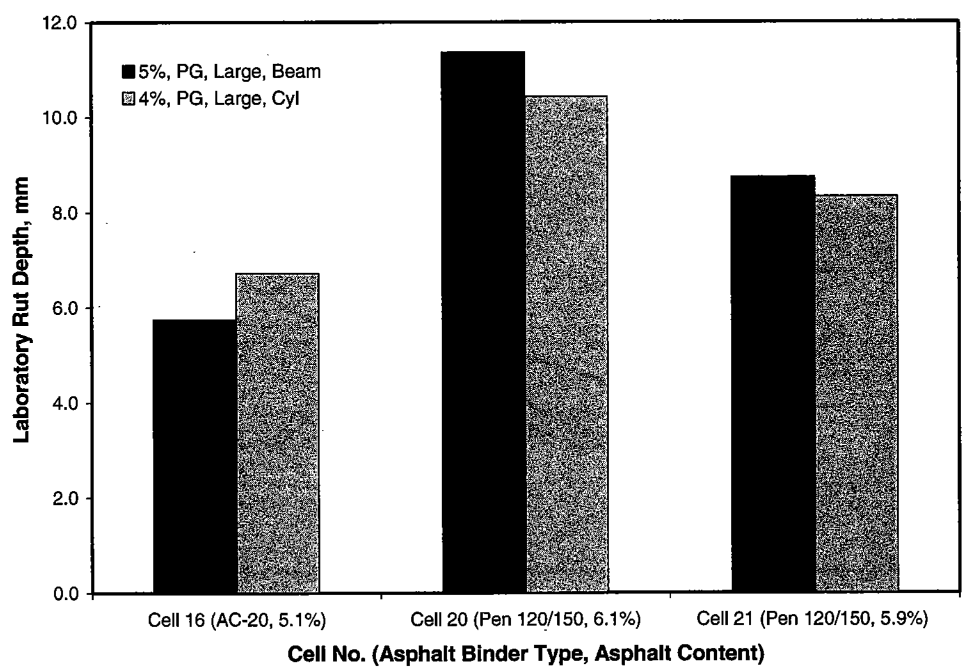


Figure 13. Comparison of MnRoad asphalt binders and asphalt contents.

TABLE 31 Average rut depths for TFHRC ALF sections (manual reading)

Air Voids	Test Temp, °C	Hose Diameter	Specimen Type	Rut Depth @ 10,000 cycles, mm			
				Lane 5	Lane 7	Lane 10	Lane 12
7%	58 (PG)	Standard	Cylinder	5.17	1.40	4.17	5.61
			Beam	9.53	2.36	5.04	7.43
		Larger	Cylinder	6.11	2.49	5.68	3.98
			Beam	7.55	2.17	4.62	4.67
	64 (PG+6)	Standard	Cylinder	6.54	2.60	5.02	8.95
			Beam	10.14	3.81	7.35	6.88
		Larger	Cylinder	6.59	3.51	5.74	6.11
			Beam	10.24	2.51	6.52	6.87
4%*	58 (PG)	Standard	Cylinder	6.47	3.19	4.84	2.38
			Beam	9.02	2.26	7.79	8.07
		Larger	Cylinder	3.71	1.59	3.68	3.15
			Beam	6.12	2.07	5.04	4.48
	64 (PG+6)	Standard	Cylinder	8.94	2.10	8.27	7.98
			Beam	7.94	3.26	8.54	9.63
		Larger	Cylinder	8.60	3.15	4.72	4.06
			Beam	10.08	2.80	5.13	6.66

* Beam samples compacted to 5.0 ± 0.5% air voids.

TABLE 32 Comparison of field and laboratory rut depth rankings for ALF

Air Voids	Test Temp., °C	Hose Diameter	Specimen Type	Field Rank[a]	Lab Rank[a]
7%	58 (PG)	Standard	Cylinder	*5, 10,* 12, 7	*12, 5, 10,* 7
			Beam	*5, 10,* 12, 7	*5, 12,* 10, 7
		Larger	Cylinder	*5, 10,* 12, 7	*5, 10,* ***12, 7*** [b]
			Beam	*5, 10,* 12, 7	5, *12, 10, 7*
	64 (PG+6)	Standard	Cylinder	*5, 10,* 12, 7	12, *5, 10,* 7
			Beam	*5, 10,* 12, 7	5, *10, 12,* 7 [b]
		Larger	Cylinder	*5, 10,* 12, 7	*5,* ***12, 10, 7***
			Beam	*5, 10,* 12, 7	5, *12, 10,* 7
4% [c]	58 (PG)	Standard	Cylinder	*5, 10,* 12, 7	5, 10, *7, 12*
			Beam	*5, 10,* 12, 7	*5, 12, 10, 7*
		Larger	Cylinder	*5, 10,* 12, 7	*5, 10,* ***12, 7*** [b]
			Beam	*5, 10,* 12, 7	*5, 10,* ***12, 7*** [b]
	64 (PG+6)	Standard	Cylinder	*5, 10,* 12, 7	*5, 10, 12,* 7 [b]
			Beam	*5, 10,* 12, 7	*12, 10, 5,* 7
		Larger	Cylinder	*5, 10,* 12, 7	5, *10, 12, 7* [b]
			Beam	*5, 10,* 12, 7	5, *12,* ***10, 7***

[a] Lanes in order of decreasing field rut depth. Italics and bold indicate statistically similar rankings. Lab rankings based upon Duncan's Multiple Range Test.
[b] Laboratory results show similar trend as field rutting.
[c] Beam samples compacted to 5.0 ± 0.5%.

combinations showing a similar trend as field performance used the low air void content.

The APA results were used along with the field rut depths for each lane to develop relationships between field and laboratory rutting. Figure 5 illustrated that Lane 5 failed after 4,000 passes with 27.4 mm of rut depth. Therefore, the rut depths of each lane after 4,000 passes of the ALF were used in this analysis. Table 33 presents the R^2 values for each of the laboratory testing combinations when compared with field rut depths. Unlike the similar WesTrack and MnRoad analyses, Table 33 does not show any negative slopes. Collectively, the R^2 values appear to be higher than the other two field experiments. All R^2 values were greater than 0.5 except one (7P+SC).

Using the data in Table 33, the six highest R^2 values were determined (Table 34). Of the six combinations with the highest R^2 values, only four had laboratory rut depths that showed a similar trend as the field rutting: 7PGLC, 5PGLB, 4PGLC, and 7P+SB. Figure 14 illustrates the relationship between field and laboratory rutting for these four combinations. All four of the testing combinations had R^2 values of 0.8885 or above, which is a strong relationship. All four combinations appear to be viable as a recommended test procedure.

The four selected ALF mixes allowed for the comparison of asphalt binder type and the effect of NMAS on laboratory rutting. Lanes 5, 7, and 10 all had identical gradations and almost identical asphalt binder contents (4.8, 4.9, and 4.9 percent, respectively). Figure 15 presents the laboratory rut depths for these three mixes using the same four testing conditions shown in Figure 14. For all four testing combinations, the mix containing the polymer-modified binder (Lane 7) had significantly lower rut depths than did the other two mixes. The polymer-modified asphalt binder was a PG 82-22. The AC-20 was a PG 64-22, and the AC-10 was a PG 58-22. Figure 15 illustrates that stiffer binders can reduce the potential for rutting.

TABLE 33 R^2 values for ALF factor-level combinations

Air Voids	Test Temp	Hose Diameter	Specimen Type	R^2
			Cylinder	0.651
		Standard	Beam	0.630
			Cylinder	0.999 [a]
	PG Temp	Larger	Beam	0.831
			Cylinder	0.373
		Standard	Beam	0.889 [a]
			Cylinder	0.774
7%	PG + 6°C Temp	Larger	Beam	0.830
			Cylinder	0.517
		Standard	Beam	0.801
			Cylinder	0.910 [a]
	PG Temp	Larger	Beam	0.917 [a]
			Cylinder	0.820 [a]
		Standard	Beam	0.634
			Cylinder	0.703 [a]
4% [b]	PG + 6°C Temp	Larger	Beam	0.662

[a] Laboratory results show similar trend as field rutting.
[b] 5% used for beams.

TABLE 34 Six highest R^2 values for ALF

Air Voids	Test Temp	Hose Diameter	Specimen Type	R^2
7%	PG	Large	Cylinder	0.999*
5%	PG	Large	Beam	0.917*
4%	PG	Large	Cylinder	0.910*
7%	PG+6	Standard	Beam	0.889*
7%	PG	Large	Beam	0.831
7%	PG	Large	Cylinder	0.830

* Laboratory results show similar trend as field rutting.

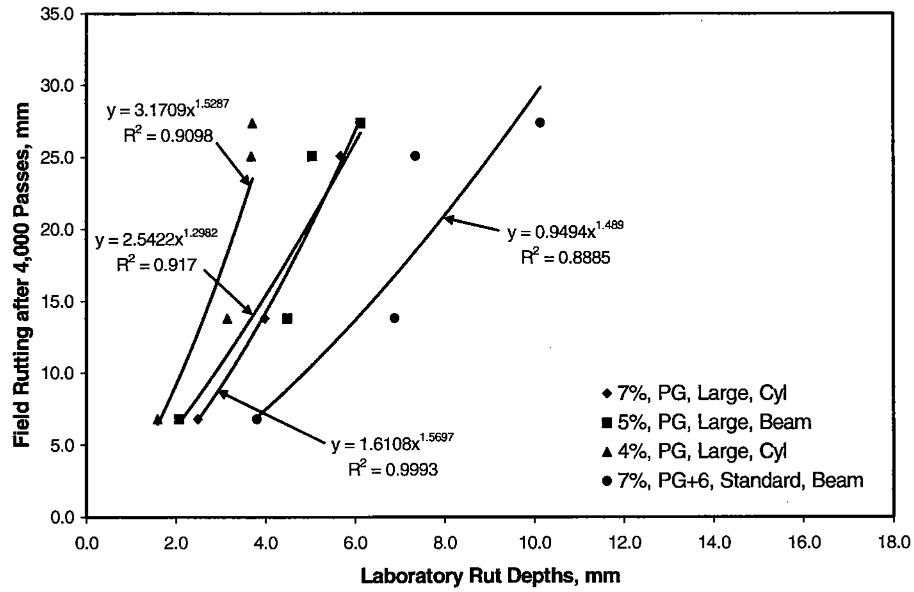


Figure 14. Four highest R^2 *values for ALF mixes.*

The effect of NMAS on laboratory rut depths (and field) can be evaluated using Lanes 10 and 12. Both of these lanes used the AC-20 binder and only differed by the NMAS and asphalt content. Figure 16 illustrates that the larger NMAS mix (Lane 12) had a lower laboratory rut depth. This also occurred in the field. Therefore, the use of larger NMAS can reduce rut susceptibility.

Another analysis that can be conducted using the ALF data is an evaluation of the effect of test temperature on mixes having polymer-modified binders. Figure 17 illustrates rut depths at the performance-grade temperature (58°C) versus rut depths at the PG+6° test temperature (64°C) for Lanes 5, 7, and 10 (all test combinations). These three lanes are shown because all three had the same gradation and differed only by

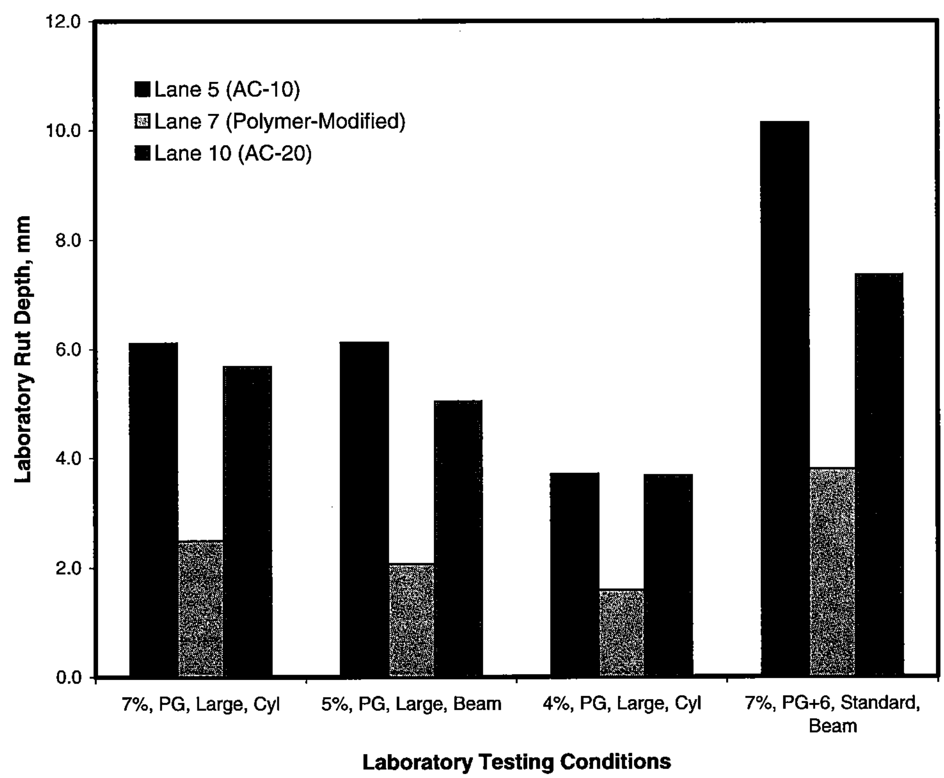


Figure 15. Comparison of neat and polymer-modified binders.

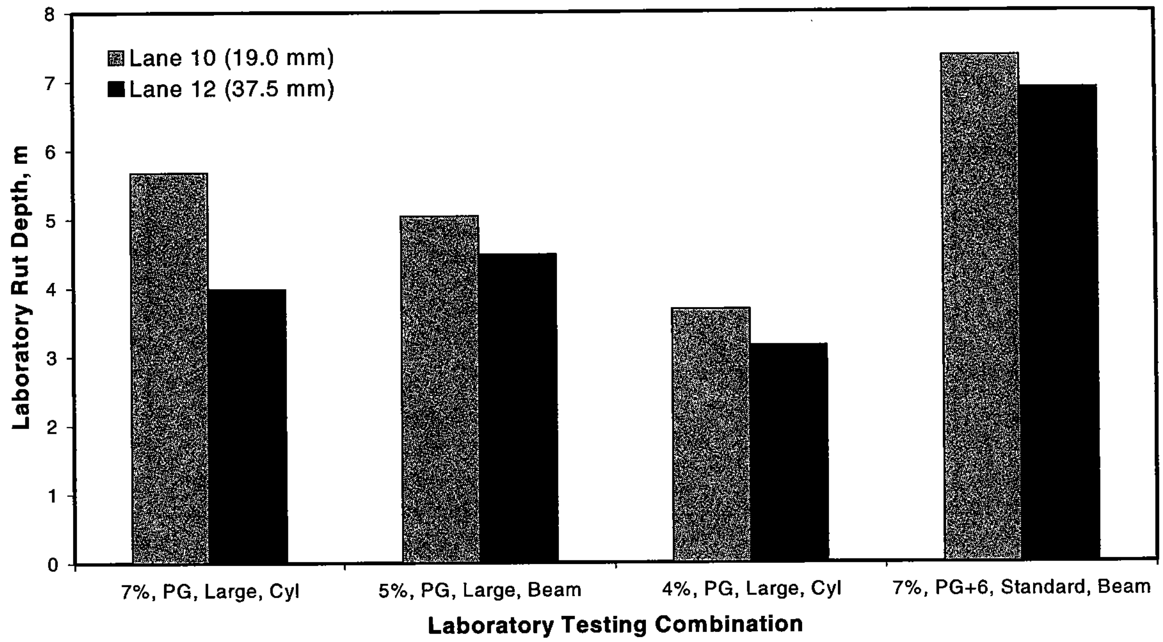


Figure 16. Effect of nominal maximum aggregate size on rutting.

asphalt binder type. This figure shows that rut depths for the three lanes had a higher magnitude when tested at 64°C than when tested at the PG temperature (58°C). As shown in Figure 15, the mix containing the polymer-modified binder had a much lower rut depth than did the AC-10 and AC-20 mixes. According to Figure 17, increasing test temperatures will result in increased rut depths, even when polymer-modified asphalt binders are used in a mix.

5.2.4 Summary of Analyses on Individual Field Projects

Table 35 summarizes the laboratory testing conditions that were most related to field results for each individual project. All of these combinations had acceptable R^2 values (above 0.835) and show similar trends to field rutting. Eight of the fourteen combinations shown in this table used beam sam-

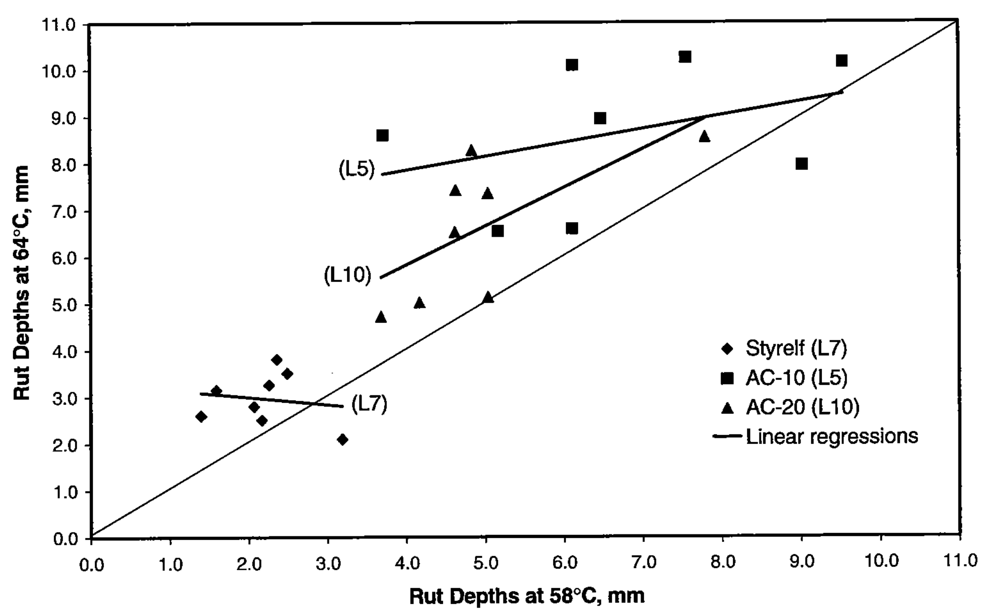


Figure 17 Effect of test temperature on polymer-modified and neat asphalts.

TABLE 35 Summary of potential testing combinations based upon analysis of individual projects

Project	Combination	R^2	Proper Trend?
WesTrack	5PGLB	1.000	Yes
	5P+SB	0.982	Yes
	4PGSC	0.856	Yes
	5PGSB	0.835	Yes
MnRoad	5PGLB	0.997	Yes
	4PGLC	0.992	Yes
	7PGLC	0.876	Yes
	7PGLB	0.863	Yes
	4PGSC	0.852	Yes
	5P+LB	0.848	Yes
ALF	7PGLC	0.999	Yes
	5PGLB	0.917	Yes
	4PGLC	0.910	Yes
	7P+SB	0.889	Yes

ples. Nine used the larger-diameter hose. Both of these factors appear to be somewhat evenly distributed. However, the test temperature and air void content factors seem to suggest one level being a more viable option. Ten of the fourteen combinations used the lower level of the air void content factor (4 percent for cylinders and 5 percent for beams). Eleven used the high temperature of the standard performance grade as the test temperature. Therefore, based only on the data generated from comparisons within individual projects, the low air void content and performance-grade test temperature appear to better predict the potential for field rutting.

5.3 ANALYSIS OF COLLECTIVE DATA (ALL PROJECTS)

This section discusses test results and analyses for the collective data. Collective data infers that the data from the three field projects were combined and analyzed.

5.3.1 Effect of Test Variables on Measured Laboratory Rut Depths

This section describes the effect of each test variable (air voids, test temperature, hose diameter, and sample type) on laboratory-measured rut depths. Figure 18 compares APA rut depths at the two levels of air void content. Surprisingly, this figure indicates that air void content does not appear to have a significant effect on measured rut depths. Previous work, through a ruggedness study with the APA (*13*), has shown air void contents to significantly affect rut depths. One potential reason for the lack of significance shown in Figure 18 is that the rut-prone mixes influenced the results. Rut depths measured in the APA include two components: consolidation of the mix and plastic flow. When comparing rut depths between 4 and 7 percent, it would be anticipated that the 4-percent samples would have lower rut depths because the consolidation component would be less. However, some of the mixes tested during the course of this study were shown to be high-rut-potential mixes (Lane 5, Lane 10, Cell 20, and Section 24). It is possible that the rut depths for the 4-percent samples included more of the plastic flow component.

Figure 19 illustrates the effect of test temperature on measured rut depths. As expected, the higher test temperature resulted in higher APA rut depths (about a 33-percent increase in rutting).

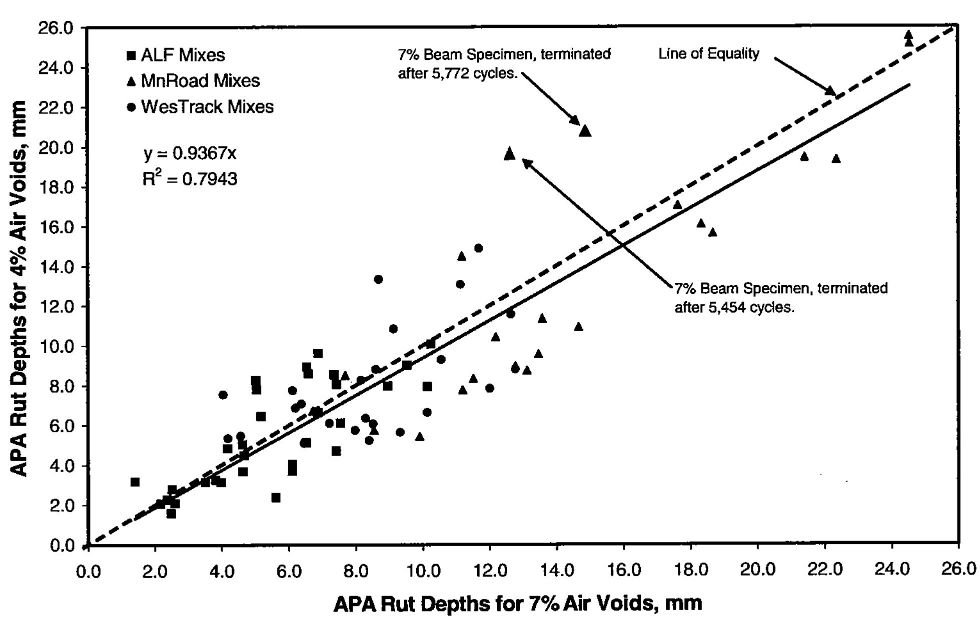


Figure 18. Effect of sample air void content on APA rut depths.

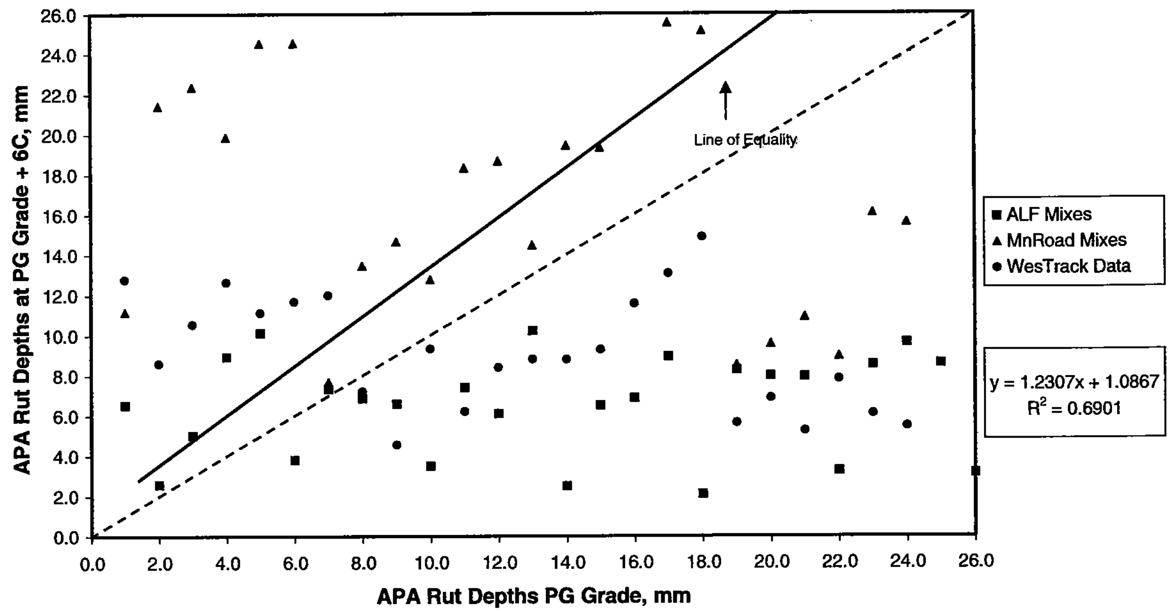


Figure 19. Effect of test temperature on APA rut depths.

Figure 20 illustrates the effect of hose diameter on APA rut depths. This factor had not been evaluated previous to the work in this study. This figure shows that the larger-diameter hose resulted in lower APA rut depths. Since the same wheel load and hose pressure were used for both hose diameters, the data appear to make sense. The larger-diameter hose distributed the wheel load over a greater area, which resulted in less stress within the contact area.

The final factor evaluated was sample type. Figure 21 shows the effect of sample type on rut depths. At low APA rut depths (<4 mm), the two sample types appear to provide similar rut depths. Above 4 mm, the beam samples provided higher rut depths. There are three potential reasons for this finding. The first reason is the mold configuration for cylindrical samples. These molds accommodate two cylinders and contain a spacer between the two samples. The spacer lies

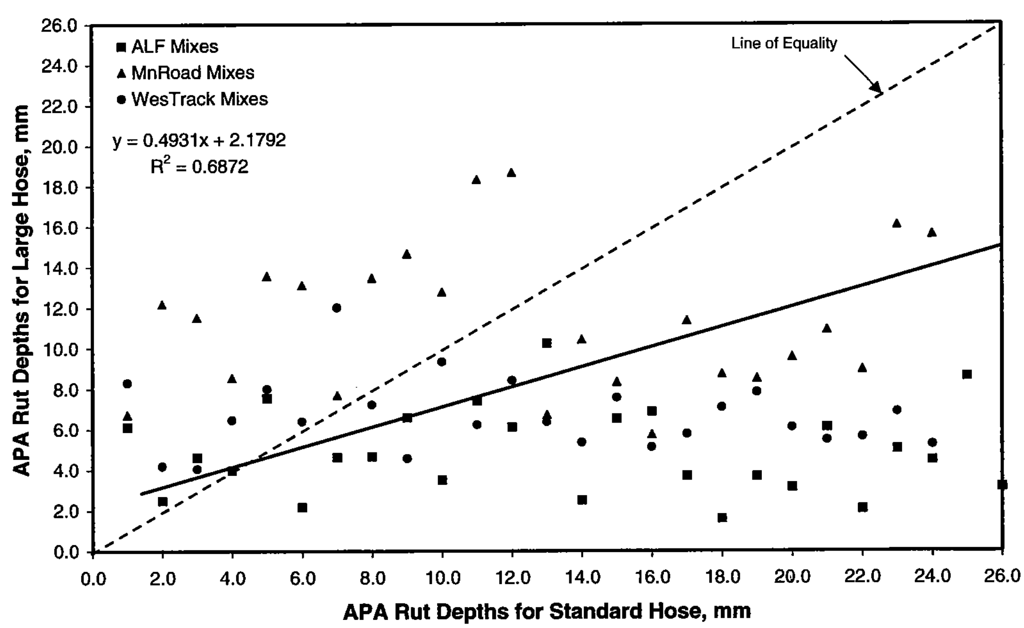


Figure 20. Effect of hose diameter on APA rut depths.

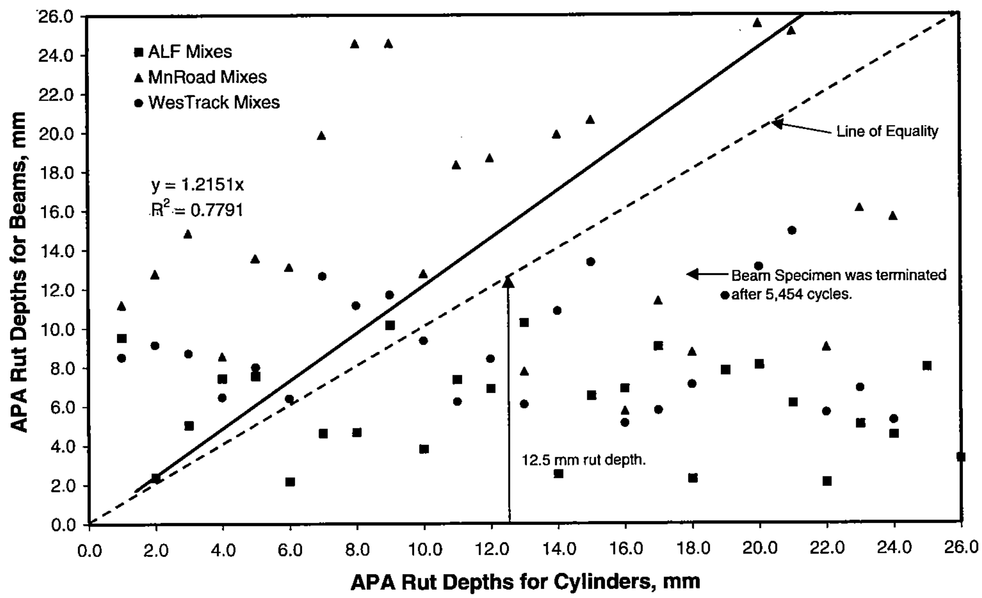


Figure 21. Effect of sample type on APA rut depths.

approximately 12.5 mm below the testing surface of specimens and is approximately 12 mm wide. For low-rut-potential mixes, the linear hose residing on the top of specimens does not approach the level of the spacer. Therefore, for low-rut-potential mixes, cylindrical specimens would not be influenced by the spacer. However, for high-rut-potential mixes, the spacer may impede the downward movement of the pressurized linear hose. This "bridging" action may be why at higher rut depths, the beam specimens yielded higher rut depths. This is shown somewhat in Figure 21 in that the beam specimens always had a higher rut depth at rut depths above about 12.5 mm.

The second potential reason for differences in rut depths between cylindrical and beam specimens is density gradients within the samples. Because the cylinders and beams were compacted using different modes of compaction, contrasting density gradients may exist in the two sample types. A study by Cooley and Kandhal (*23*) showed that the two sample types do have different density gradients, and this may also explain the differences between the two.

The third potential reason could be the slightly higher air void contents in the lower level of the air void content factor. Beam samples were compacted to 5 ± 0.5 percent while the cylinders were compacted to 4 ± 0.5 percent.

5.3.2 Comparison of Field and Laboratory Rutting

Within this section, data from the three field experiments (WesTrack, MnRoad, and ALF) were collectively analyzed in order to evaluate the different testing conditions. Analyses included developing the relationship (R^2 values) between field rut depths and laboratory rut depths. Because each of the test pavements (sections, cells, and lanes) had a different volume of traffic, a method of normalizing the traffic levels on each test pavement was needed. The method selected was to divide actual field rut depths by the square root of the number of ESALs experienced at the time of rut depth measurements. This method has been used previously by Brown and Cross (*1*).

Table 36 presents the field rut depths and ESAL values used during the analyses conducted with the collective data. The rut depth values shown in this table are slightly different than those used in the previous analyses on the individual projects (in Section 5.2). For the collective comparisons, it was decided that final cumulative rut depths and ESALs were more desirable. ESAL values for the ALF project shown in the table represent one pass of the ALF. As stated in Chapter 3, the assembly employed by the ALF equated to one-half of an 18-kip axle.

At first, an attempt was made to correlate field rut depths to APA laboratory rut depths using data from all three projects. However, a negative trend was observed in all of the correlations as shown by a typical plot (Figure 22). The likely reason for the negative trends when all three projects were "lumped" together was the traffic characterization from the ALF pavements. One pass of the ALF was not equivalent to one pass of an 18-kip axle. Therefore, the MnRoad and WesTrack data were combined in order to evaluate the combined relationships between field and laboratory rutting.

Table 37 presents the R^2 values for each of the laboratory testing combinations when MnRoad and WesTrack field rut depths were compared with laboratory rutting. R^2 values having a negative sign indicates that the relationship between field and laboratory rut depths was opposite that which would be expected because the regression line showed decreasing labo-

TABLE 36 Rut depths and ESALs used in analysis of collective data

Project	Mix	Field Rut Depth, mm	ESALs	Rut Depth / SQRT(ESAL)
	Section 15	9.2	5,003,303	0.00411
	Section 19	14.5	5,003,303	0.00648
WesTrack	Section 24	26.0	2,774,052	0.01561
	Cell 16	5.0	3,051,267	0.00286
	Cell 20	18.8	2,423,667	0.01208
MnRoad	Cell 21	12.1	3,051,267	0.00693
	Lane 5	27.4	4,000	0.43323
	Lane 7	18.1	200,000	0.04047
	Lane 10	36.3	10,000	0.36300
ALF*	Lane 12	24.1	200,000	0.05389

*ALF ESAL counts are equal to 1 pass of the ALF.

ratory rut depths as field rutting increased. Of the 16 laboratory testing conditions, only 2 showed laboratory rut depths that followed the same trend as field rutting for both the MnRoad and WesTrack mixes (4PGSC and 5PGLB). All of the R^2 values are 0.501 or below, which indicates that the relationships are not strong. Four combinations have R^2 values greater than 0.28, as shown in Table 38, and represent the four most viable options for a tentative test procedure. All four of these combinations used the low level of the air void content factor and the standard temperature of the performance grade as the test temperature. This was not surprising because most of the combinations that correctly ranked the individual projects also used the lower air void contents and performance-grade test temperature (see Table 35). Two of the combinations used the standard hose, and two used the larger-diameter hose. The two highest R^2 values shown in Table 38 used beam samples, while the other two used cylinders.

Figures 23 through 26 illustrate the relationships of the four laboratory testing combinations shown in Table 38. Included in each figure are error bars for each data point that represent plus or minus one standard deviation for the three replicates per mixture. The three replicates (two cylinders or one beam per replicate) within a test were used to calculate the standard deviations.

Figure 23 illustrates the relationship between field and laboratory rutting (with MnRoad and WesTrack data combined) for the 4PGLC combination. The regression line for this relationship shows the expected trend: laboratory rut depths increased as the field rutting rate increased; however, the R^2 for the relationship was very low (0.28). An analysis of vari-

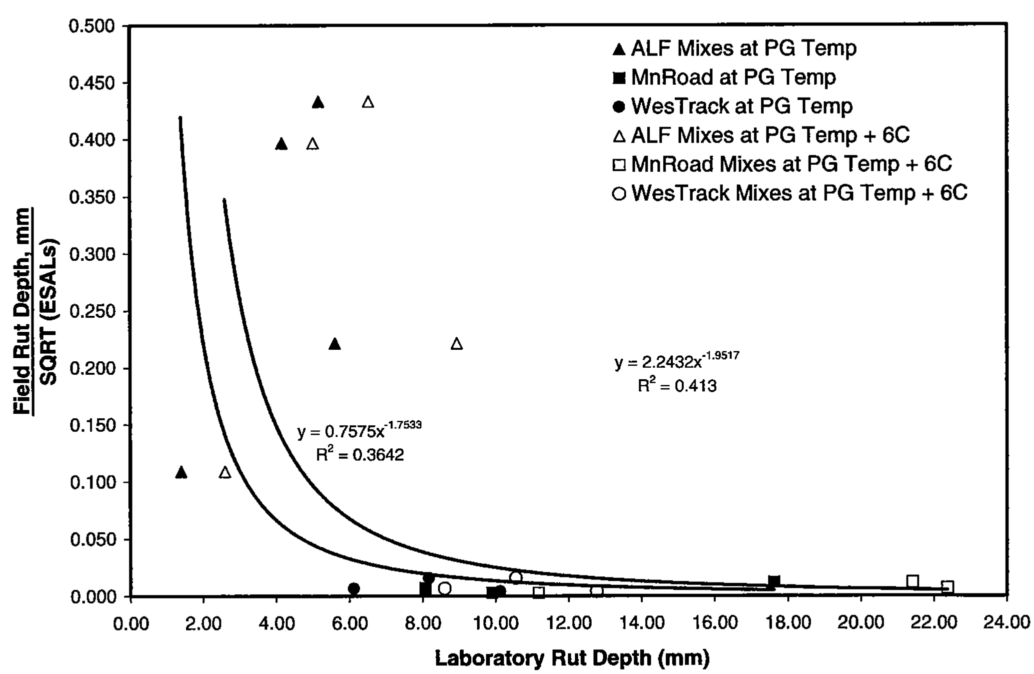


Figure 22. Cylindrical specimens, 7% air voids, standard hose.

TABLE 37 R^2 values for MnRoad and WesTrack data combined

Air Voids	Test Temp, °C	Hose Diameter	Specimen Type	$R^{2\,a}$
			Cylinder	0.029
		Standard	Beam	0.001
			Cylinder	–0.010
	60 (PG)	Larger	Beam	0.028
			Cylinder	0.036
		Standard	Beam	–0.002
			Cylinder	–0.051
7%	70 (PG + 6)	Larger	Beam	0.002
			Cylinder	0.365[b]
		Standard	Beam	0.501
			Cylinder	0.282
	64 (PG)	Larger	Beam	0.428[b]
			Cylinder	0.005
		Standard	Beam	–0.011
			Cylinder	–0.079
4%[c]	70 (PG + 6)	Larger	Beam	0.009

[a] A negative sign indicates that the relationship between field and laboratory rutting was opposite of what would be considered acceptable.
[b] Laboratory results show the same trend as field rutting for both projects.
[c] Beam samples compacted to 5.0±0.5% air voids.

TABLE 38 Six highest R^2 values for MnRoad and WesTrack data combined

Air Voids	Test Temp	Hose Diameter	Specimen Type	R^2
5%	PG	Standard	Beam	0.501
5%	PG	Large	Beam	0.428*
4%	PG	Standard	Cylinder	0.365*
4%	PG	Large	Cylinder	0.282

*Laboratory results show the same trend as field rutting for both projects.

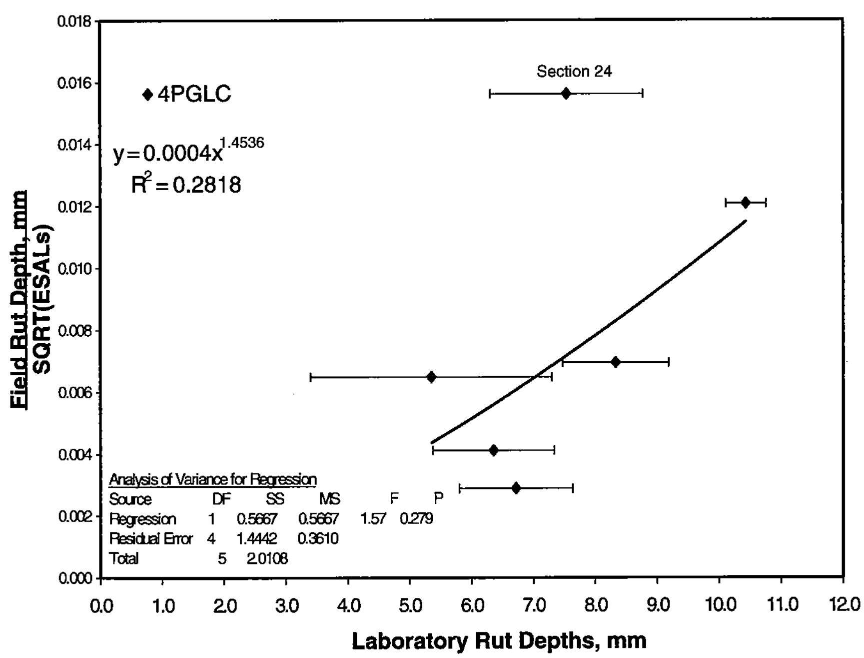


Figure 23. Combined WesTrack and MnRoad data for 4PGLC combination.

ance (ANOVA) for the regression (also shown in the figure) indicated that the relationship was not significant at a 5-percent level of significance ($\alpha = 0.05$). Also, based upon the data shown in Figure 23, it appears that the results from the WesTrack Section 24 may be outliers.

The relationship between field and laboratory rutting for the 5PGLB combination is illustrated in Figure 24. Again, the regression line shows the expected trend. The R^2 for the relationship is stronger for the 5PGLB combination (0.43) than for the 4PGLC combination (0.28). However, an ANOVA indicated that the relationship was not significant because the probability of the F-statistic being greater than F-critical (P-value) was greater than 0.05 (0.159). Results shown in Figure 24 again suggest that the results from Section 24 may be outliers.

Figure 25 illustrates the relationship between field and laboratory rutting for the 4PGSC combination. The R^2 for this relationship is not strong at 0.3647. The insignificance of the relationship was confirmed by an ANOVA (also shown in the

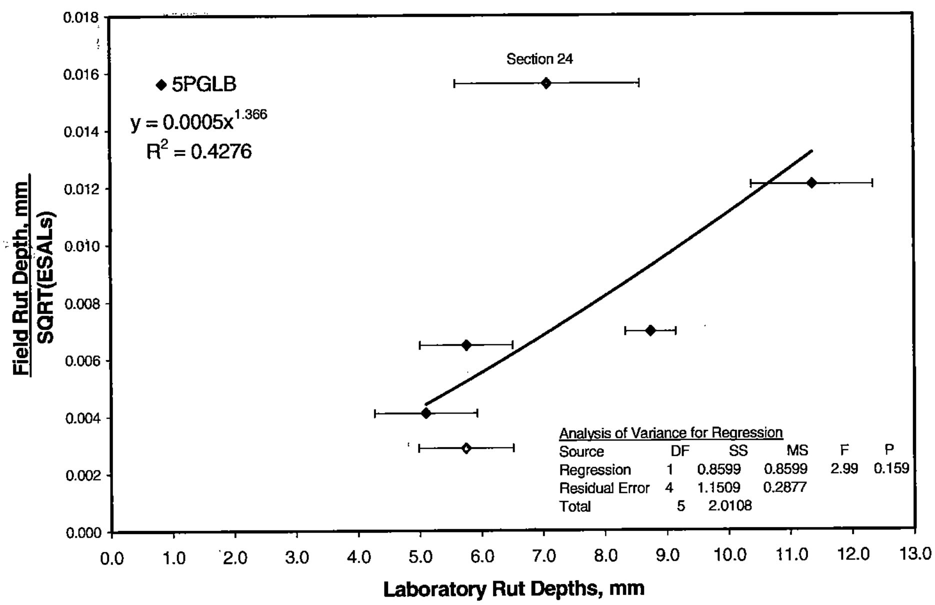


Figure 24. Combined WesTrack and MnRoad data for 5PGLB combination.

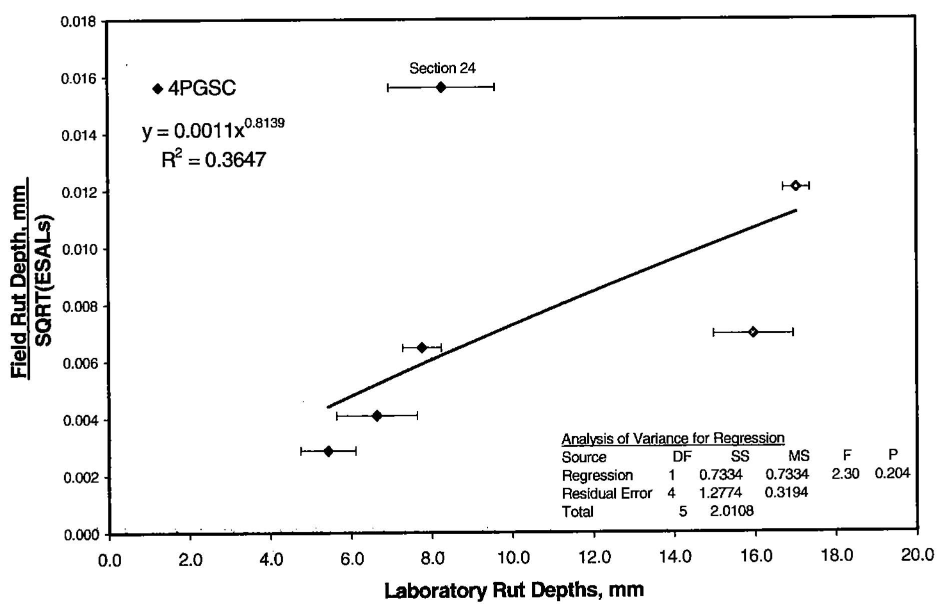


Figure 25. Combined WesTrack and MnRoad data for 4PGSC combination.

figure) as the P-value was greater than 0.05. Similar to the 4PGLC and 5PGLB combination results, the results from Section 24 appear to be greatly influencing the R^2 value and may be outliers.

The relationship between field and laboratory rutting for the 5PGSB combination is illustrated in Figure 26. Of the four combinations evaluated, the 5PGSB combination had the highest R^2 value at 0.50. However, an ANOVA showed that the relationship was not significant (a P-value of 0.115). Again, the results from Section 24 occur the farthest from the trend line and may be outliers.

Because the results from WesTrack Section 24 may have been outlying data for each of the four combinations, the Section 24 data were removed from the data set and regression analyses were conducted again. Figure 27 illustrates the 4PGLC combination with the Section 24 data removed. The

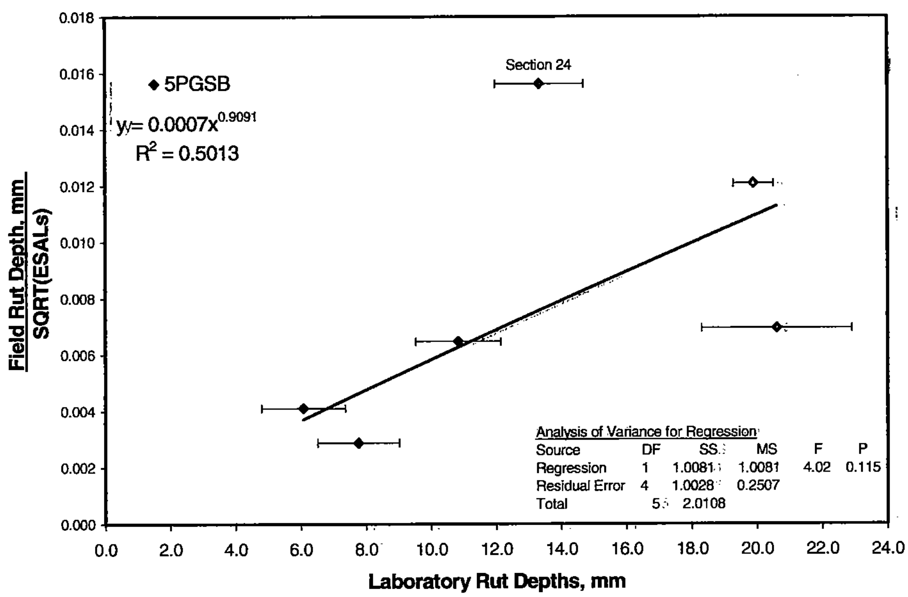


Figure 26. Combined WesTrack and MnRoad data for 5PGSB combination.

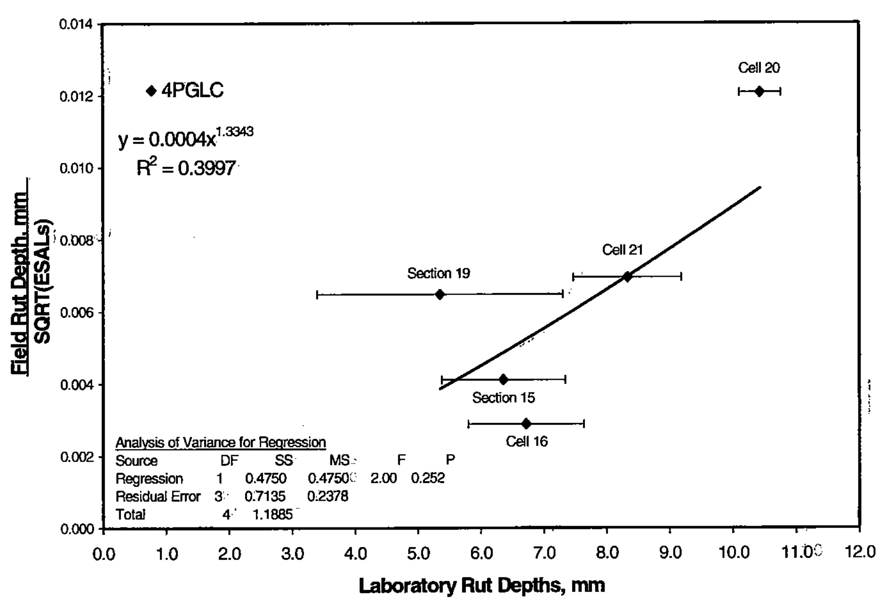


Figure 27. Combined WesTrack and MnRoad data for 4PGLC combination, excluding WesTrack Section 24.

R^2 value improved with the removal of the Section 24 data. With Section 24 data included, the R^2 value was 0.28, and without the Section 24 data the R^2 value was 0.40. However, an ANOVA showed that the removal of the Section 24 data did not result in a significant relationship between field and laboratory rutting (a *P*-value of 0.252).

Figure 28 illustrates the 5PGLB combination with Section 24 results excluded. Again, the R^2 value increased with the exclusion of the Section 24 data (0.43 to 0.70). An ANOVA indicated that the relationship was not significant at a level of significance of 5 percent (a *P*-value of 0.077); however, the relationship was significant at a 10-percent level of significance (a *P*-value of less than 0.10).

The relationship between field and laboratory rutting with Section 24 removed for the 4PGSC combination is illustrated in Figure 29. With the removal of Section 24, the R^2 value

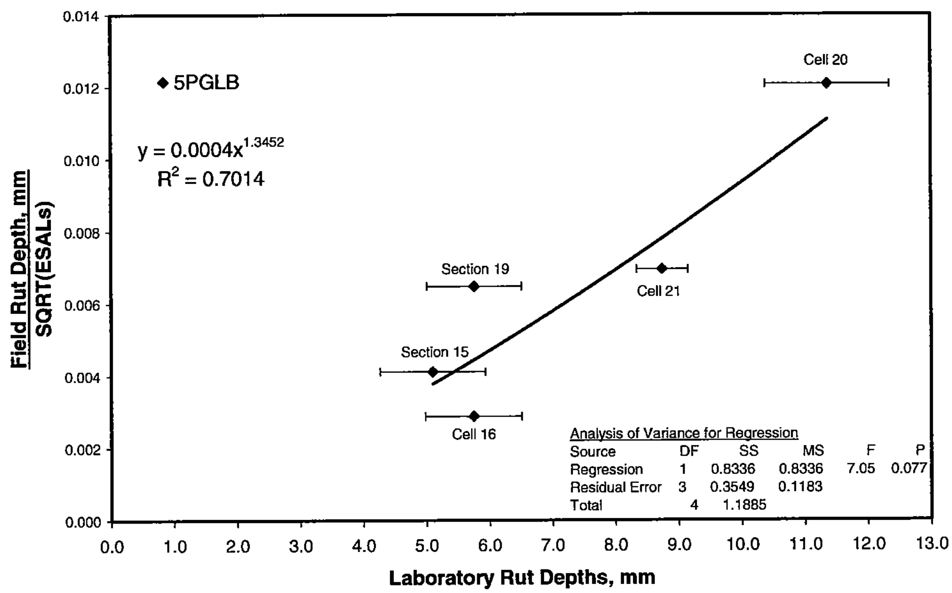


Figure 28. Combined WesTrack and MnRoad data for 5PGLB combination, excluding WesTrack Section 24.

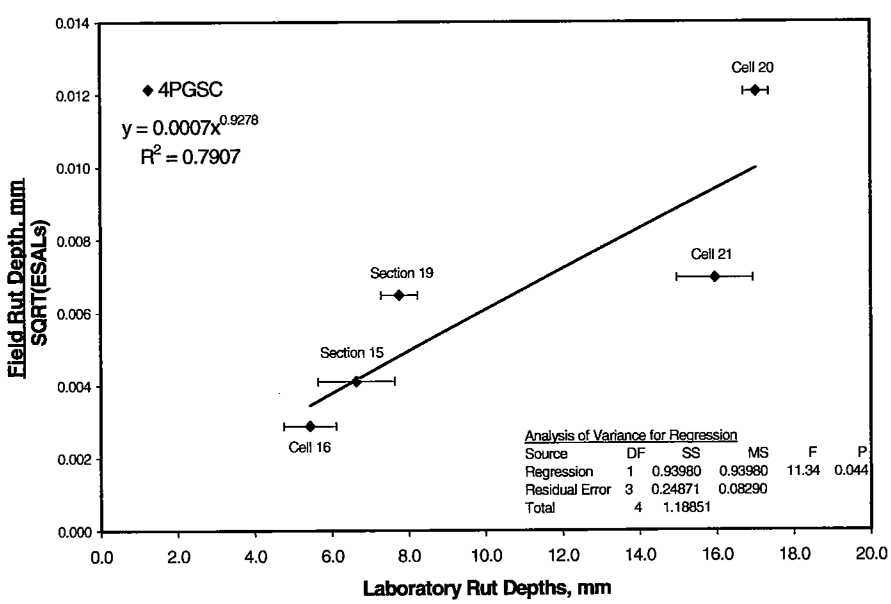


Figure 29. Combined WesTrack and MnRoad data for 4PGSC combination, excluding WesTrack Section 24.

increased for this combination from 0.36 to 0.79. An ANOVA confirmed that the relationship was significant at a 5-percent level of significance (a P-value of 0.044).

Figure 30 illustrates the 5PGSB combination with the Section 24 results removed. As with the other three combinations, the R^2 value increased (from 0.50 to 0.69) when Section 24 results were removed. An ANOVA indicated that the relationship was still not significant at a 5-percent level of significance; however, the relationship was significant at a level of significance of 10 percent (a P-value of 0.081).

An interesting observation about the comparisons with and without the WesTrack Section 24 was that the removal of the Section 24 data had little effect on the best-fitted line. For all four combinations, the regression equations are similar.

Prior to making a selection as to the best laboratory test combination(s), it was necessary to look at the variability within each of the test combinations. Two separate analyses were conducted to evaluate the variability within each of the four test procedures. First, using data from each of the 10 field mixes, the standard deviation for a given mix was plotted versus the average laboratory rut depth. The slope of this line is the COV. Figures 31 and 32 present the COV plots for the combinations containing cylindrical and beam specimens, respectively. COVs determined from both plots indicate less

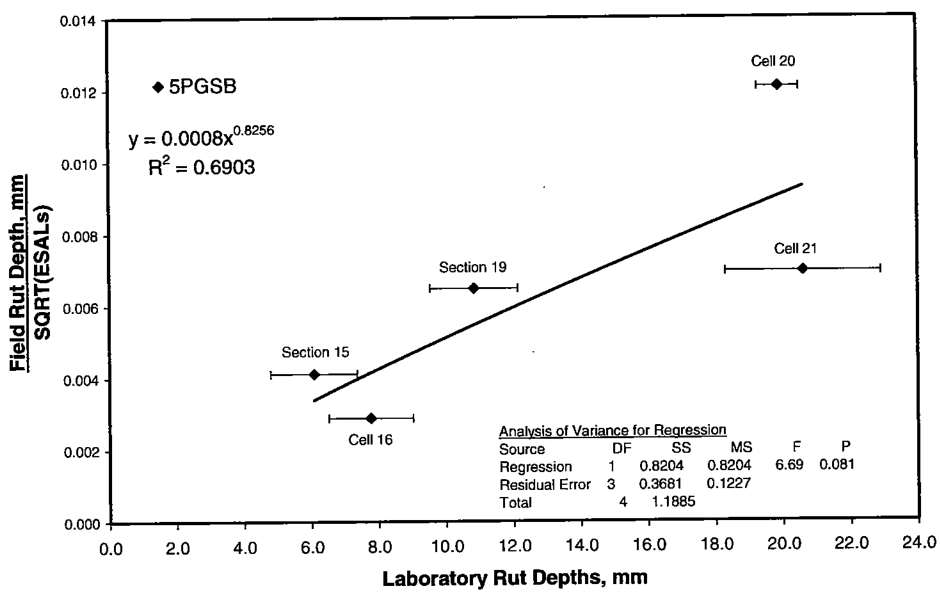


Figure 30. Combined WesTrack and MnRoad data for 5PGSB combination, excluding WesTrack Section 24.

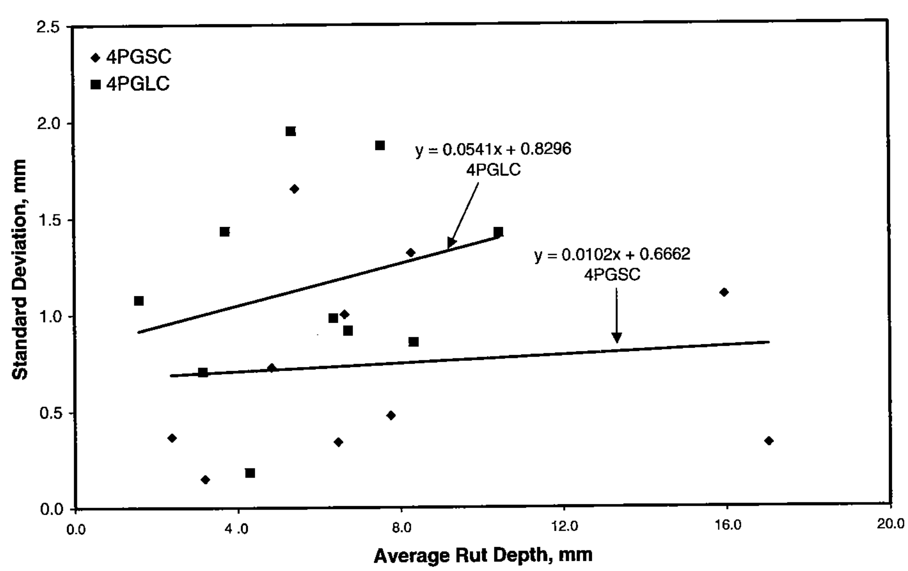


Figure 31. Comparison of testing combination variability (cylindrical samples).

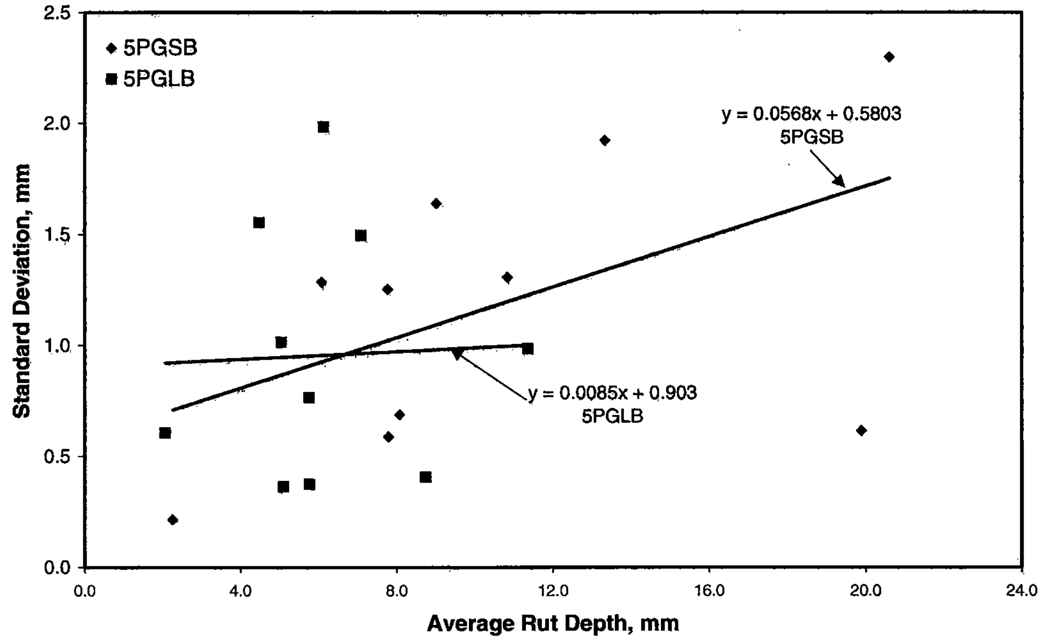


Figure 32. Comparison of testing combination variability (beam samples).

than 6-percent variation. The COVs for the two combinations containing beams were 5.7 and 0.9 percent. The COVs for the two combinations containing cylinders were 5.4 and 1.0 percent. All four of these COVs are considered acceptable; therefore, either sample type or hose diameter could be recommended in the tentative test procedure.

In order to evaluate whether the standard or larger-diameter hose showed less variability, six cylindrical samples compacted to 4-percent air voids were made for each of the 10 mixes. Half of these 10 mixes were randomly selected to be tested at the performance-grade temperature and standard hose, and the other half were tested at the performance-grade temperature and larger hose. These tests were intended as replicate tests (instead of replicates within a test) so that variability between tests (or repeatability) could be evaluated. Table 39 presents the results of both replicate tests for each mix. Also included in this table is a column labeled "Residuals." The residuals are based on a line of equality between the first and second replicates. This line of equality would have the form "$y = x$." Analysis of the data in Table 39 involved determining the average squared residual. This property is calculated by squaring each of the residuals, summing the squared residuals for a given combination, and dividing by $n - 2$ (where $n = 5$). As the average squared residual increases, the variation about the line of equality also increases. For the 4PGLC combinations, the average squared residual was 6.86;

TABLE 39 Results of replicate tests on 4-percent cylindrical samples

Combination	Mix	Replicate 1	Replicate 2	Residuals
	Lane 7	1.89	1.59	0.30
	Cell 16	6.03	6.72	–0.69
	Cell 20	7.07	10.43	–3.36
	Cell 21	6.22	8.33	–2.11
	Section 15	4.29	6.36	–2.07
4PGLC			*Average Squared Residual*	*6.86*
	Lane 5	6.47	6.07	0.40
	Lane 10	4.84	5.56	–0.72
	Lane 12	2.38	4.15	–1.77
	Section 19	7.76	6.23	1.53
	Section 24	8.27	6.12	2.15
4PGSC			*Average Squared Residual*	*3.59*

for the 4PGSC combinations, the average squared residual was 3.59. This indicates that the standard hose combinations had less variability about the line of equality.

5.4 SELECTION OF TENTATIVE TEST PROCEDURE AND ACCEPTANCE CRITERIA

The primary objectives of Phase II of this study were to determine the suitability of the APA as a general method of predicting the rut potential of HMA mixes and to develop a tentative standard procedure for the APA. Based upon the analyses of individual and combined data, four potential test combinations were identified: 4PGSC, 4PGLC, 5PGSB, and 5PGLB. An evaluation of the COV for each combination indicated that all four had acceptable within-test variabilities as all had COV values of less than 6 percent. Another analysis compared the reproducibility of test results for the two hose diameters evaluated using cylindrical samples compacted to 4-percent air voids and tested at the high temperature of the standard performance grade. This analysis indicated that tests with the standard-diameter hose were more repeatable. In addition to this analysis on hose diameter, it was shown that test combinations containing the standard-diameter hose had more amplification between well- and poorly performing mixes. This was shown by the relative flatness of the regression lines for the standard-hose combinations compared with the larger-hose combinations. Therefore, three of the four factors evaluated in Phase II appear to be justifiable: low air void content, test temperature corresponding to the high temperature of the standard performance grade, and the standard-diameter hose. Because no definitive justification for either beam or cylindrical samples could be reached, both sample types are recommended for inclusion in the tentative test procedure.

Another part of the overall objective for this study was to develop preliminary critical rut depths that could be used as acceptance criteria. Since two APA testing combinations were selected, criteria were developed for both. The method of selecting critical rut depths was based on the relationship between field and laboratory rut depths when the MnRoad and WesTrack sections were combined. Figures 33 and 34 illustrate this relationship for the 4PGSC and 5PGSB combinations, respectively, and represent the same relationships shown in Figures 25 and 26. The method of determining the tentative acceptance criteria can be explained in Figure 33. First, a critical field rut depth of 12.5 mm is assumed. This field rut depth represents the maximum acceptable level of rutting in the field. Second, the y-axis of Figure 33 is the field rut depths from the MnRoad and WesTrack projects divided by the square root of the ESALs. The ESALs in this figure represent the number of ESALs obtained at the time field rut depths were measured. Using the assumed critical value of field rutting (12.5 mm) and a given traffic (ESAL) level allowed for the acceptance criteria to be taken from Figures 33 and 34. The traffic levels selected were 2, 3, 5, 10, and 30 million ESALs. Table 40 presents the values used to determine the laboratory acceptance criteria.

Based upon Figures 33 and 34, Table 41 presents acceptance criteria recommended for the two selected APA test procedures based on manual rut depth measurements after 10,000

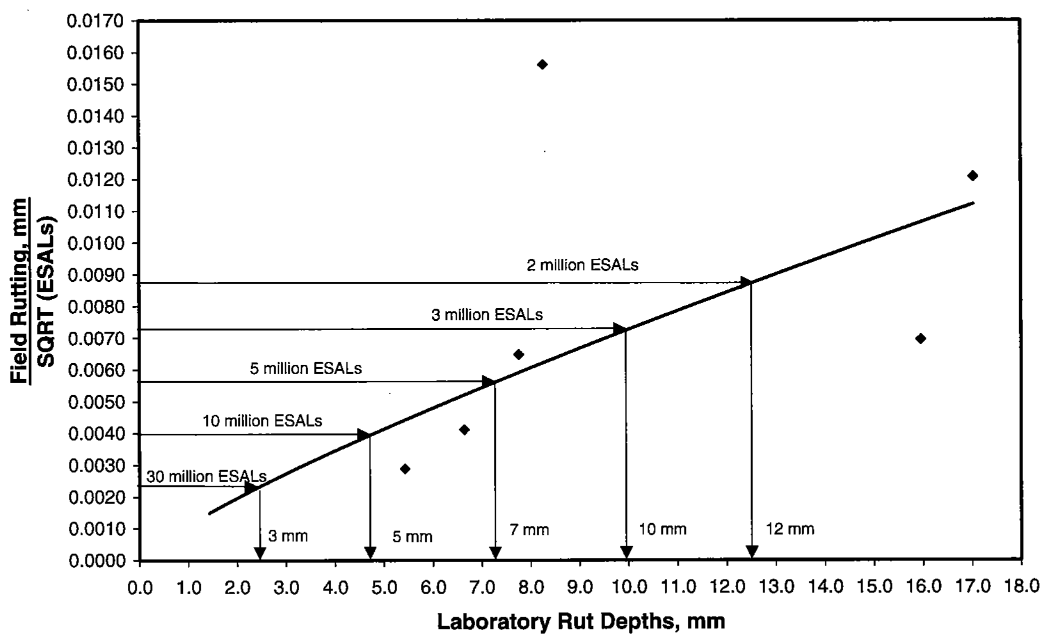


Figure 33. Selected combination 4PGSC sample.

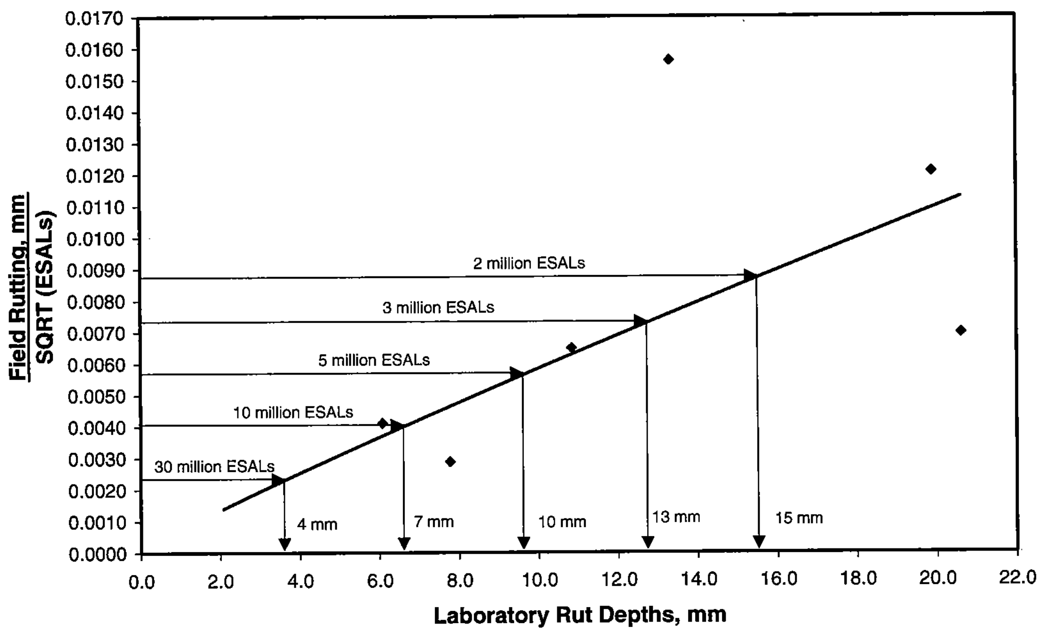


Figure 34. Selected combination 5PGSB sample.

cycles. The acceptance criteria appear logical. As the traffic level increases, the required rut resistance also increases. The 4PGSC acceptance criteria are slightly lower than the 5PGSB criteria; however, recall that beam samples provided higher rut depths during the experiment and that the beams had slightly higher air voids (5 versus 4 percent).

A number of APAs are equipped with automatic rut depth measuring systems. Therefore, a comparison between manual and automatic measurements after 10,000 cycles was conducted. Figure 35 presents this comparison, which indicates a strong relationship ($R^2 = 0.97$). This figure shows that the manual measurements consistently provided higher rut depths and that the difference was not a constant offset. Using the regression line shown in Figure 35, tentative acceptance criteria were developed for automatic rut depth measurements and are presented in Table 42. Values in this table have been rounded to the nearest 0.5 mm.

Based upon a recent APA User's Group meeting, all agencies that use an APA specification have criteria based upon 8,000 cycles. Therefore, a similar comparison was conducted between manual measurements after 10,000 cycles and automatic measurements after 8,000 cycles (see Figure 36). Again, a strong relationship was observed ($R^2 = 0.96$). Table 43 presents tentative acceptance criteria for automatic rut depth measurements after 8,000 cycles (rounded to nearest 0.5 mm).

Because manual rut depth measurements were only obtained after 10,000 cycles, a comparison to obtain a set of tentative acceptance criteria for manual measurements after 8,000 cycles cannot be conducted. However, if the assumption that the relationship between manual and automatic rut depth measurements shown in Figure 35 is also viable at 8,000 cycles,

TABLE 40 Assumed rut depth and ESAL values for acceptance criteria

Critical Rut Depth, mm	ESALs	Field Rut Depth, mm / SQRT (ESALs)
12.5	2,000,000	0.0088
12.5	3,000,000	0.0072
12.5	5,000,000	0.0056
12.5	10,000,000	0.0040
12.5	30,000,000	0.0023

TABLE 41 Tentative APA acceptance criteria: manual measurements after 10,000 cycles

Traffic Level, ESALs	Combination	
	4PGSC[1]	5PGSB[2]
2 million	12 mm	15 mm
3 million	10 mm	13 mm
5 million	7 mm	10 mm
10 million	5 mm	7 mm
30 million	3 mm	4 mm

[1] 4-percent air voids, PG test temperature, standard hose, cylindrical sample.
[2] 5-percent air voids, PG test temperature, standard hose, beam sample.

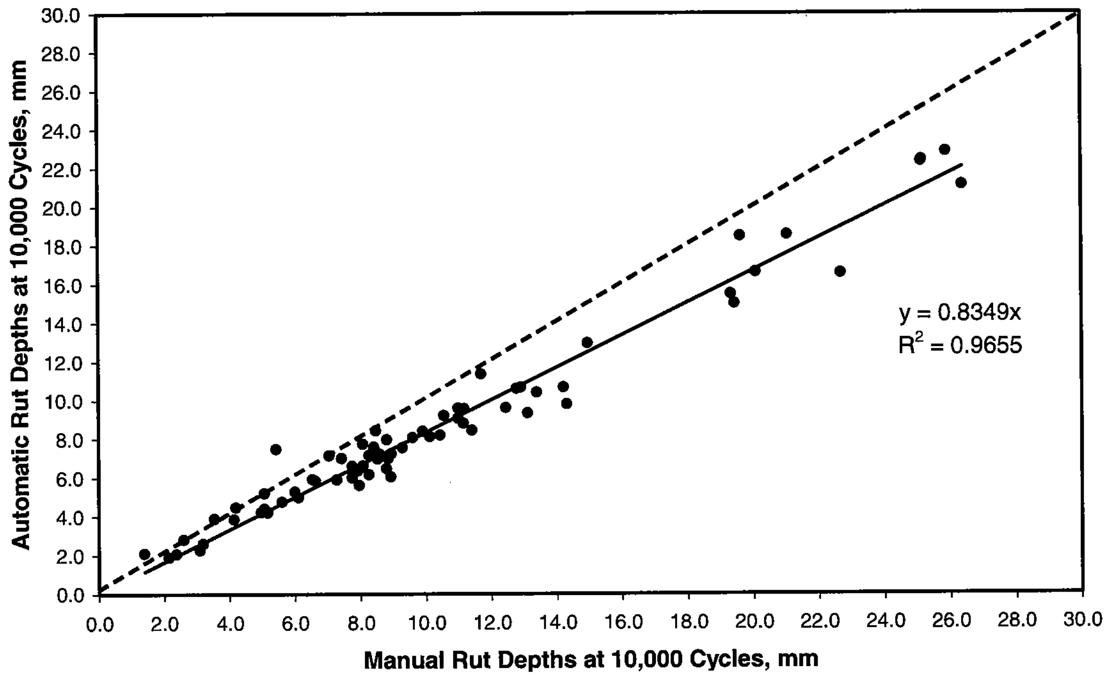


Figure 35. Comparison of manual (10,000 cycles) and automatic (10,000 cycles) rut depths.

the information in Table 43 can be used to provide acceptance criteria for manual measurements after 8,000 cycles (see Table 44). Again, values in this table were rounded to the nearest 0.5 mm.

One other comparison was conducted between manual and automatic rut depth measurements. This comparison was between manual measurements after 10,000 cycles and automatic measurements after 4,000 cycles (see Figure 37). This comparison was made in anticipation of recommending a test plan on evaluating the APA for quality control testing. Figure 37 illustrates that there was a reasonably strong relationship between the 10,000 and 4,000 cycle data (R^2 = 0.87). This indicates that during quality control testing, the APA may be able to differentiate between a well- and poorly performing mix after 4,000 cycles.

TABLE 42 Tentative APA acceptance criteria: automatic measurements after 10,000 cycles

	Combination[1]	
Traffic Level, ESALs	4PGSC[2]	5PGSB[3]
2 million	10.0 mm	12.5 mm
3 million	8.5 mm	11.0 mm
5 million	6.0 mm	8.5 mm
10 million	4.0 mm	6.0 mm
30 million	2.5 mm	3.5 mm

[1] Values have been rounded to the nearest 0.5 mm.
[2] 4-percent air voids, PG test temperature, standard hose, cylindrical sample.
[3] 5-percent air voids, PG test temperature, standard hose, beam sample.

5.5 COMPARISON BETWEEN APA RESULTS AND OTHER PERFORMANCE TESTS

Selected mixes from the Phase II research were used to compare the relationship between field and laboratory rutting for the APA, HWTD, and simple performance test being developed as part of NCHRP Project 9-19.

5.5.1 Hamburg Wheel Tracking Device

HWTD testing was carried out at APAC Materials Services in Atlanta, Georgia. Only mixes from the ALF and MnRoad field experiments were tested because limited materials were available from the WesTrack field experiment.

The test parameters used with the HWTD were a target air void content of 6 ± 0.5 percent, a test temperature of 55°C, and a wheel load of 667 N (150 lb). Table 45 presents the results of the HWTD testing conducted on the ALF and MnRoad mixes. As shown within the table, two replicates were tested for each mix. One replicate test of MnRoad Cell 21 failed prior to achieving the end of the consolidation period. Also, the two replicates for the MnRoad Cell 20 had a wide variation although both suggest high potential for rutting.

Table 46 presents the average APA rutting results for the two recommended APA test configurations (4PGSC and 5PGSB) for the same mixes shown in Table 45 and the aver-

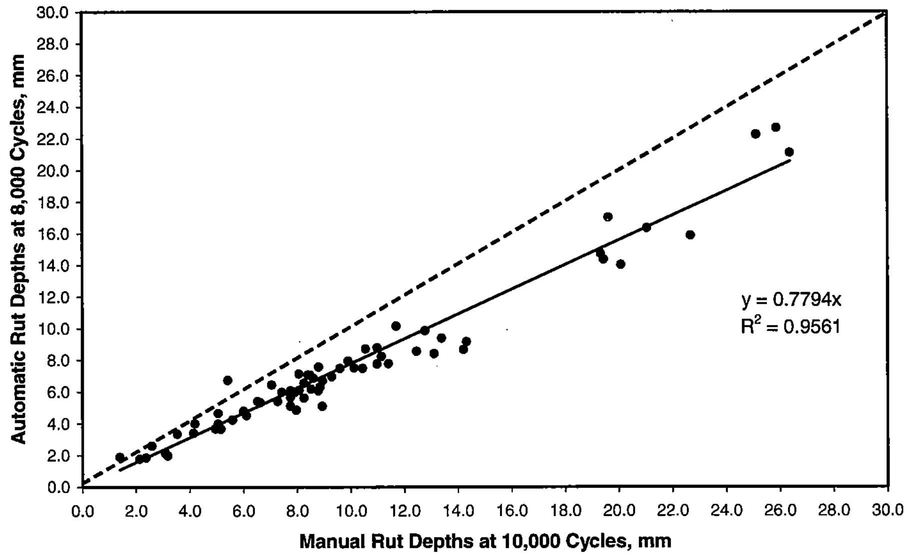


Figure 36. Comparison of manual (10,000 cycles) and automatic (8,000 cycles) rut depths.

TABLE 43 Tentative APA acceptance criteria: automatic measurements after 8,000 cycles

	Combination[1]	
Traffic Level, ESALs	4PGSC [2]	5PGSB [3]
2 million	9.5 mm	11.5 mm
3 million	8.0 mm	10.0 mm
5 million	5.5 mm	7.5 mm
10 million	4.0 mm	5.5 mm
30 million	2.5 mm	3.0 mm

[1] Values have been rounded to the nearest 0.5 mm.
[2] 4-percent air voids, PG test temperature, standard hose, cylindrical sample.
[3] 5-percent air voids, PG test temperature, standard hose, beam sample.

TABLE 44 Tentative APA acceptance criteria: manual measurements after 8,000 cycles

	Combination [1]	
Traffic Level, ESALs	4PGSC [2]	5PGSB [3]
2 million	11.0 mm	14.0 mm
3 million	9.5 mm	12.0 mm
5 million	6.5 mm	9.5 mm
10 million	4.5 mm	6.5 mm
30 million	3.0 mm	3.5 mm

[1] Values have been rounded to the nearest 0.5 mm.
[2] 4-percent air voids, PG test temperature, standard hose, cylindrical sample.
[3] 5-percent air voids, PG test temperature, standard hose, beam sample.

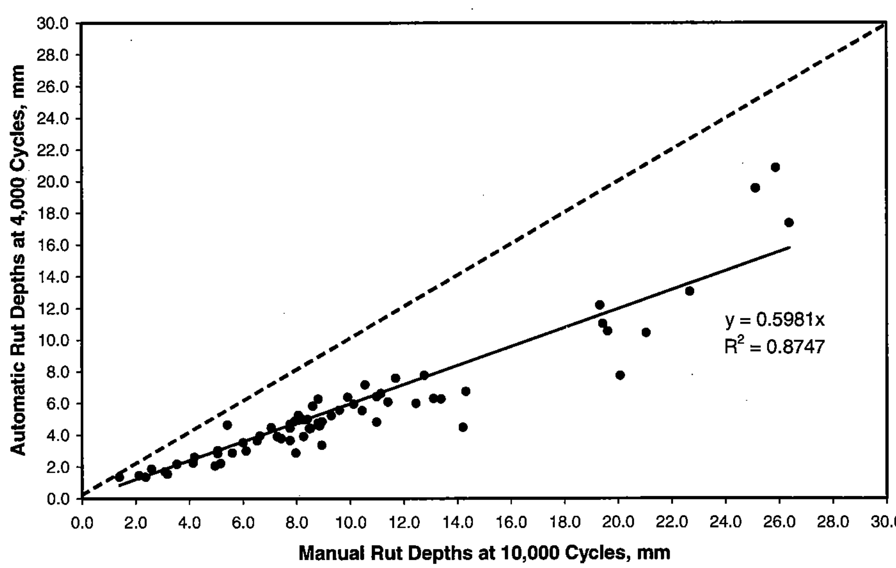


Figure 37. Comparison of manual (10,000 cycles) and automatic (4,000 cycles) rut depths.

TABLE 45 Results of Hamburg WTD testing on ALF and MnRoad mixes (mm/hr)

Field Experiment	Mixture	Replicate 1	Replicate 2	Average
	Lane 5	4.1	7.3	5.7
	Lane 7	0.2	0.3	0.3
	Lane 10	4.9	3.9	4.4
FHWA ALF	Lane 12	1.4	3.3	2.4
	Cell 16	4.4	1.6	3.0
	Cell 20	114.8	41.0	77.9
MnRoad	Cell 21	53.4	[a]	[b]

[a] Sample failed prior to reaching end of consolidation. No average calculated.
[b] No average determined.

TABLE 46 Results of APA and Hamburg WTD testing on ALF and MnRoad mixes (mm/hr)

Field Experiment	Mixture	4PGSC, mm[a]	5PGSB, mm[b]	HWTD (mm/hr)
	Lane 5	6.47	9.02	5.7
	Lane 7	3.19	2.26	0.3
	Lane 10	4.84	7.79	4.4
FHWA ALF	Lane 12	2.38	8.07	2.4
	Cell 16	5.43	7.77	3.0
	Cell 20	17.03	19.89	77.9
MnRoad	Cell 21	15.96	20.61	[c]

[a] APA cylindrical samples at 4-percent air voids and tested at the standard PG temperature.
[b] APA beam samples at 5-percent air voids and tested at the standard PG temperature.
[c] Sample failed prior to reaching end of consolidation. No average calculated.

age HWTD rutting rates. All three sets of data appear to show a similar trend. For the ALF mixes, Lane 7 provided low rut depths (and rutting rate) and Lane 5 showed the highest potential for rutting. For the MnRoad mixes, Cell 16 showed the lowest potential for rutting for all three data sets while Cells 20 and 21 both showed very high potential for rutting. The results for all three test combinations shown in Table 46 are illustrated in Figure 38. Not included in Figure 38 are the results from Cell 21 because one of the replicates did not complete consolidation rutting during HWTD testing.

Figure 38 illustrates that the test results from both APA testing procedures and the HWTD are related. At APA rut depths below approximately 10 mm, the relationship between APA and HWTD results appears to be linear. For Cell 20, both the APA and HWTD suggest a very high potential for rutting. Figure 39 illustrates the relationship between laboratory rutting (rut depth or rutting rate) and field rutting at the ALF. For this analysis, the field rut depths after 4,000 ALF passes were used. All three test procedures show the expected trend in the data. As field rutting increases, laboratory rutting also increases. The regression lines for all three test procedures also appear to have somewhat similar slopes between field and laboratory rutting. Results from the HWTD had the highest coefficient of correlation (R^2) at 0.95, followed by the 5PGSB ($R^2 = 0.80$) and 4PGSC ($R^2 = 0.52$) APA combinations, respectively.

Figure 40 compares the results of HWTD testing and APA testing for the two recommended procedures with results of field rutting on the MnRoad project. Field rut depths on this figure were after 2 million ESALs. No regression lines were generated for this figure because of the lack of HWTD results on Cell 21. As expected, for the three mixes the APA results are generally similar. However, the range of test results between Cell 16 and Cell 20 for the HWTD is much greater. The HWTD results confirm that Cell 20 had a high potential for rutting.

5.5.2 Simple Performance Test

A further comparison was made between the APA and the candidate simple performance tests for rutting identified in

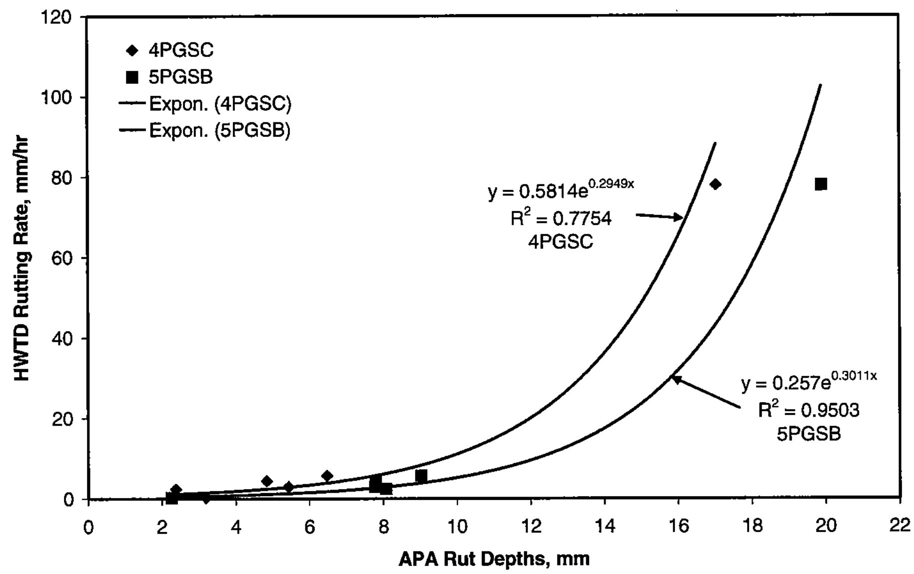


Figure 38. Comparison of APA and HWTD rutting results for ALF and MnRoad mixes.

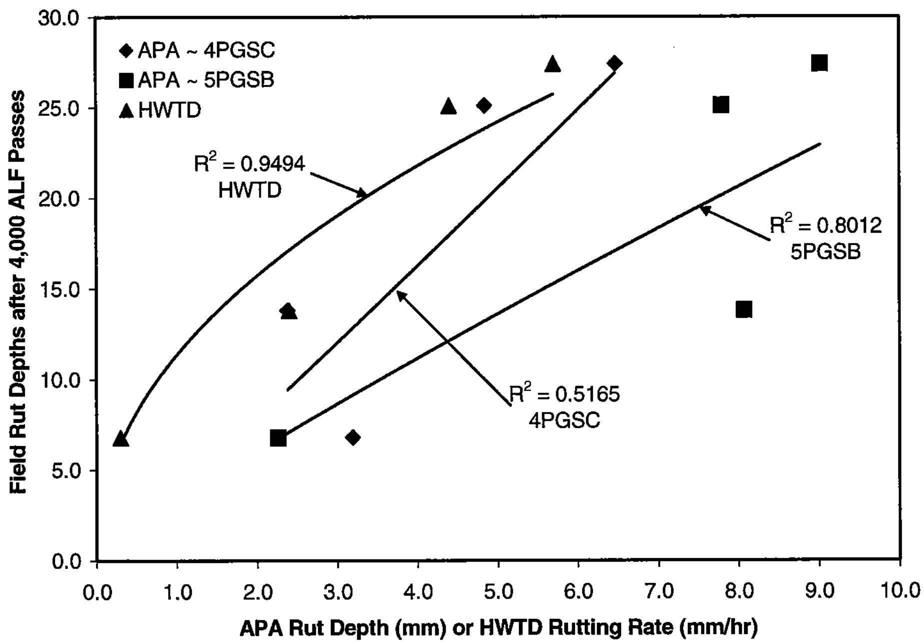


Figure 39. Comparison of HWTD and two recommended APA procedures to field rutting for the ALF field experiment.

NCHRP Project 9-19: the dynamic modulus test ($E^*/\sin\delta$); the flow time as measured by the triaxial creep test (F_T); and the flow number as measured by a repeated load triaxial test (F_N).

The research agency contacted the research team for NCHRP Project 9-19 and obtained test results for the three simple performance tests (SPTs) in which similar mixes were tested. Mixes for which both SPT and APA results were available included MnRoad Cell 16, MnRoad Cell 20, ALF Lane 5, ALF Lane 7, ALF Lane 10, ALF Lane 12, WesTrack Section 15, and WesTrack Section 24. SPT results for these sections are presented in Table 47.

Initial analysis of the data within Table 47 compared the results of APA testing with each of the SPT results. Figures 41 through 43 provide the relationship between the APA results and $E^*/\sin\delta$, F_T, and F_N, respectively. These relationships have R^2 values ranging from 0.26 to 0.63. The strongest

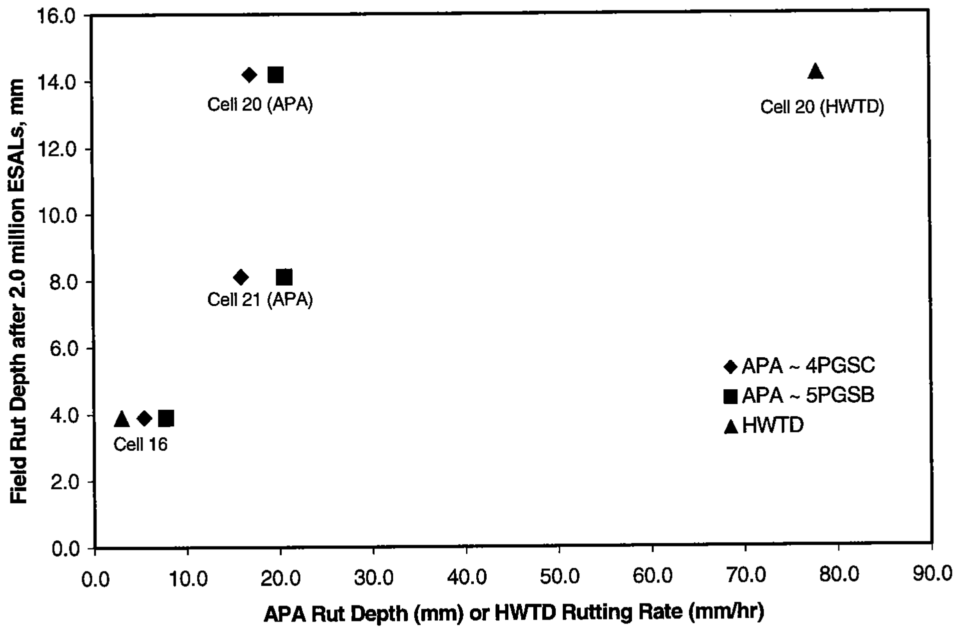


Figure 40. Comparison of HWTD and two recommended APA procedures to field rutting for MnRoad field experiment.

TABLE 47 Results of APA and SPT testing on similar mixes

Field Experiment	Mixture	APA Combinations		Simple Performance Tests		
		4PGSC	5PGSB	$E^*/\sin\delta$, 10^6 psi [a]	Flow Time, sec [b]	Flow Number, repetitions [c]
	Lane 5	6.47	9.02	0.0959	8933	349
	Lane 7	3.19	2.26	0.1757	None [d]	7761
	Lane 10	4.84	7.79	0.0756	11765	1658
FHWA ALF	Lane 12	2.38	8.07	0.1472	11290	None [d]
	Cell 16	5.43	7.77	0.1020	1141	444
MnRoad	Cell 20	17.03	19.89	0.0625	77	144
	Section 15	6.64	6.08	0.1344	[e]	[e]
WesTrack	Section 24	8.27	13.33	0.1341	121970	1756

[a] Dynamic Modulus Test ~ Test temperature of 130°F and frequency of 5 Hz.
[b] Static Creep Test ~ Test temperature of 130°F, confining stress of 20 psi and axial stress of 140 psi for ALF and WesTrack and 120 psi for MnRoad.
[c] Unconfined Repeated Load Triaxial Test ~ Test temperature of 130°F and axial stress of 10 psi.
[d] None—no flow within the 10,000 cycles specified by the test procedure.
[e] WesTrack Section 15 was determined as an outlier for all tests under NCHRP Project 9-19.

relationship for both APA test configurations was with flow number (see Figure 43). The poorest relationship for both APA test configurations was with the flow time (see Figure 42). Figure 42 shows that the Section 24 data appear to be outliers for the flow time results.

The next analysis with the SPT data was to compare laboratory APA and SPT results with field rut depths. This analysis was similar to the analysis conducted earlier in this report using normalized rut depths; the ALF data were not used because the ALF wheel passes could not be related to ESALs.

Results of the comparisons between $E^*/\sin\delta$, F_T, and F_N and normalized field rutting are illustrated in Figures 44 through 46, respectively. The relationships between APA rut depths and normalized field rut depths are also shown on all three figures. For the comparisons, the axes were reversed so that a second y-axis (represeting the SPT results) could be added. This allowed a better visual comparison of the different test procedures. Four mixes are included on these figures: MnRoad Cell 16, MnRoad Cell 20, WesTrack Section 15, and WesTrack Section 24. Flow time and flow number results for the WesTrack Section 15 mix-

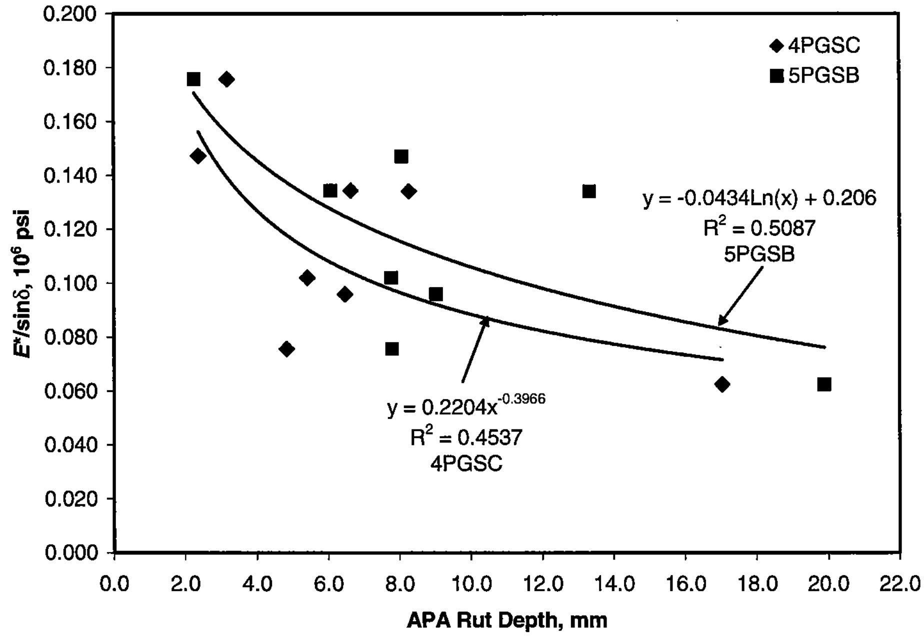


Figure 41. Comparison of E*/sinδ *with APA results.*

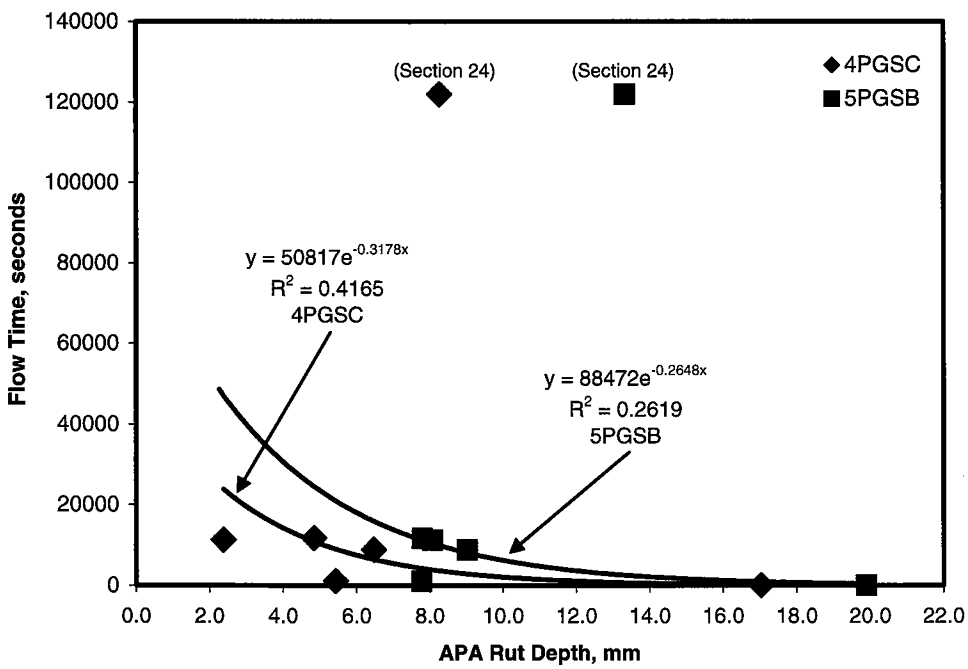


Figure 42. Comparison of flow time (F_T) *with APA results.*

ture were considered outliers by the NCHRP Project 9-19 researchers and therefore were not provided.

Figure 44 illustrates the two recommended APA test procedures and $E^*/\sin\delta$ results versus normalized field rut depths. The R^2 values on this figure range from 0.05 to 0.70 with the 5PGSB having the largest R^2 value and $E^*/\sin\delta$ having the lowest. The regression line for $E^*/\sin\delta$ has a negative slope. This indicates that as stiffness decreased, the potential for rutting increased.

Figure 45 illustrates the relationship between the two APA procedures and flow time versus normalized field rutting. The R^2 values for the three regression lines range from 0.50 to 0.70 with the 5PGSB having the highest and flow time having the lowest R^2 value. As stated previously, the flow

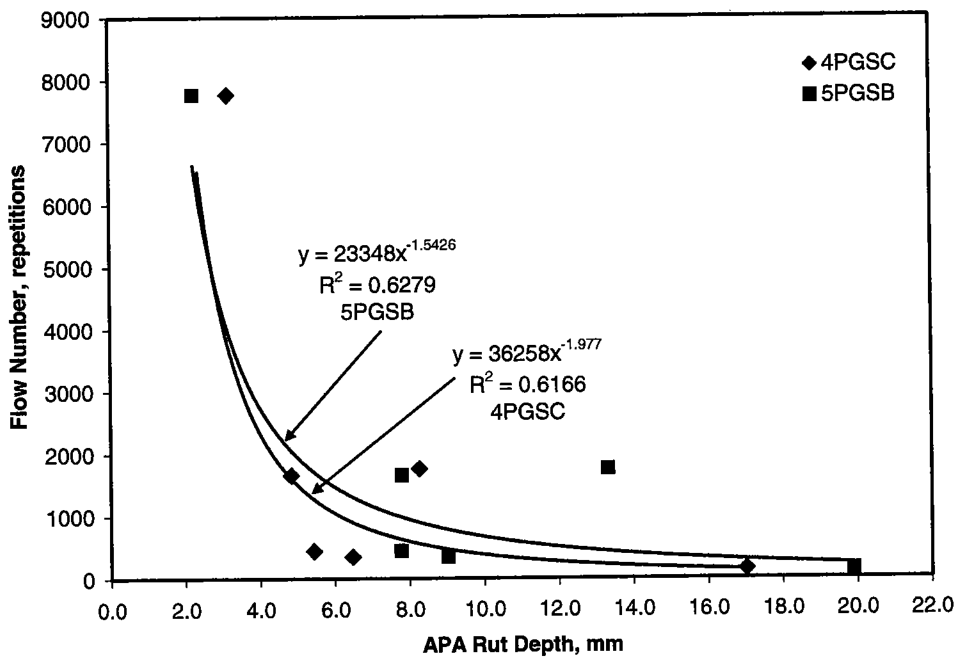


Figure 43. Comparison of flow number (F_N) *with APA results.*

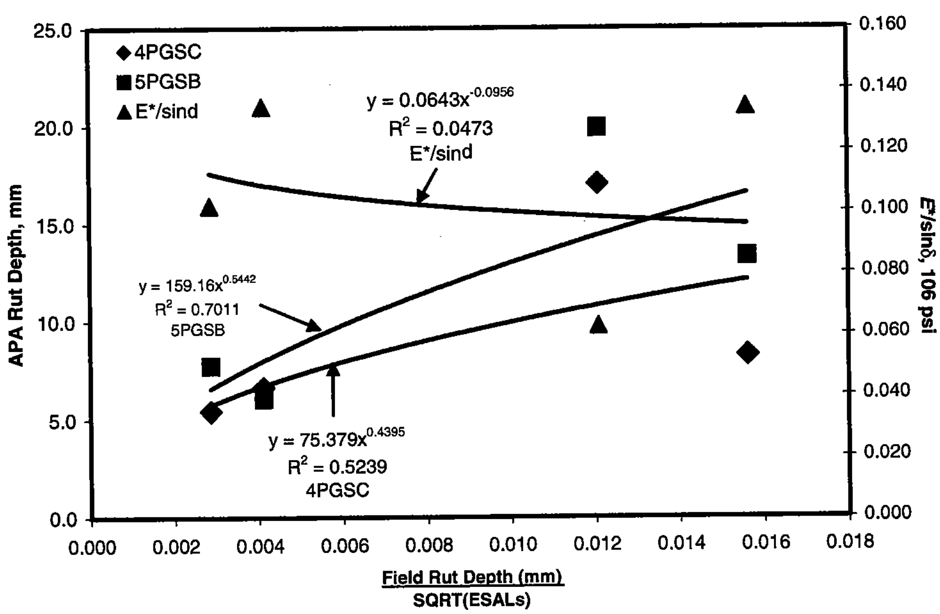


Figure 44. Comparison of recommended APA procedures and E*/sinδ *with field rutting.*

time data do not include WesTrack Section 15 and, therefore, contains only three data points compared with the four data points for the APA data. All three regression lines have a somewhat similar slope.

Figure 46 illustrates the two recommended APA test procedures and the flow number results versus normalized field rut depths. Similar to the flow time analysis, the flow number data shown in Figure 46 contains only three data points. The R^2 values on this figure range from 0.34 to 0.70 with the 5PGSB having the largest R^2 value and flow number having the lowest. All three regression lines have somewhat similar slopes.

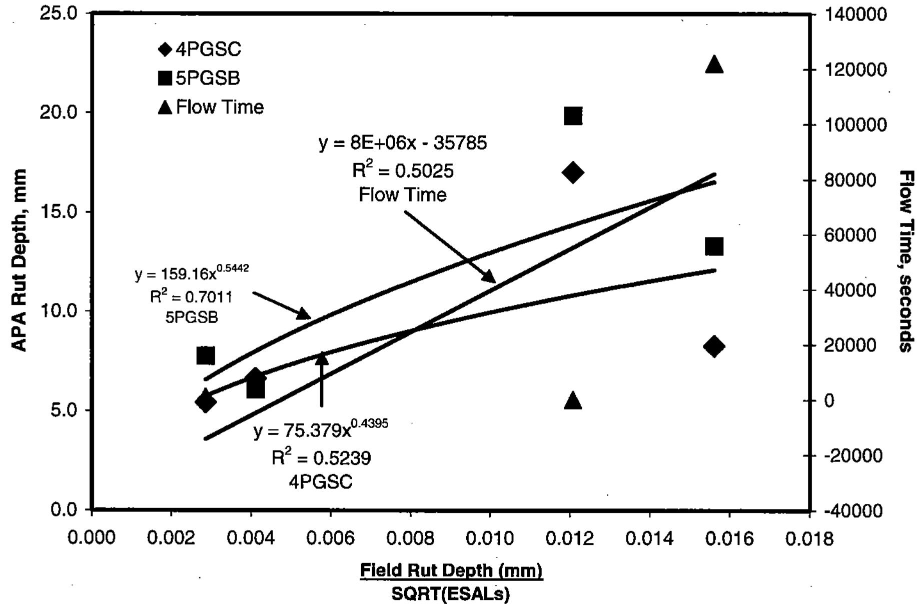


Figure 45. Comparison of recommended APA test procedures and flow time (F_T) *with field rutting.*

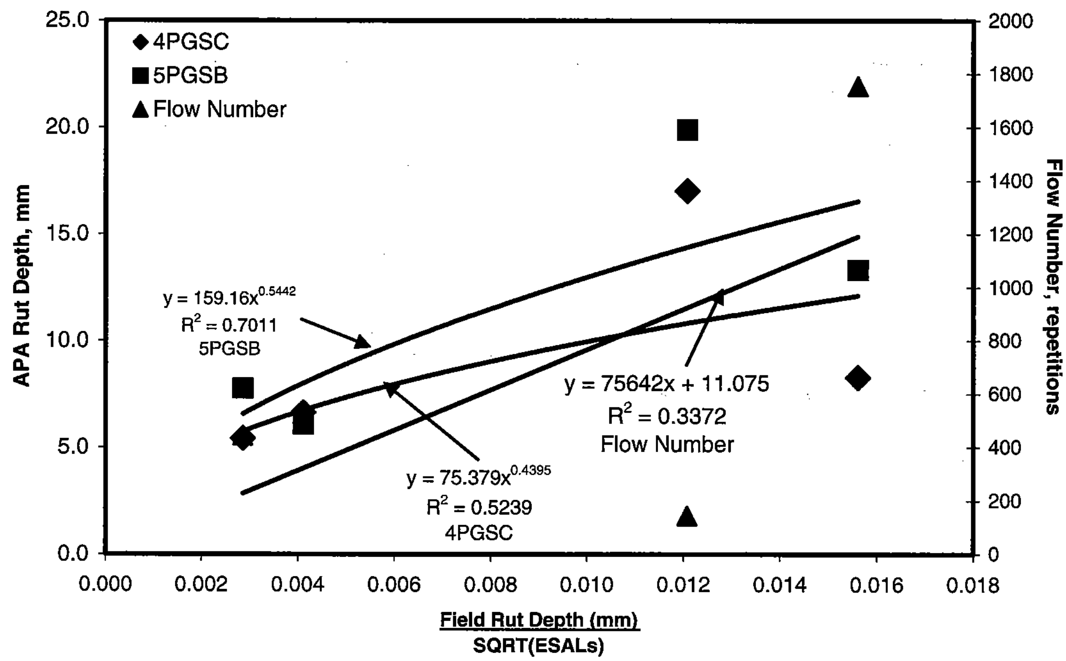


Figure 46. Comparison of recommended APA test procedures and flow number (F_N) *with field rutting.*

CHAPTER 6

PHASE II CONCLUSIONS AND RECOMMENDATIONS

6.1 CONCLUSIONS

The conclusions provided below are based upon the research results obtained during Phase II:

1. Cylindrical samples compacted to 4-percent air voids and beam samples compacted to 5-percent air voids resulted in APA laboratory test results that were more closely related to field rutting performance than were cylindrical and beam samples compacted to 7-percent air voids.
2. Samples tested in the APA at a test temperature corresponding to the high temperature of the standard performance grade for a project location better predicted field rutting performance than did samples tested at 6°C higher than the high temperature of the standard performance grade.
3. Samples with both the standard and large-diameter hoses predicted field rutting performance about equally. However, samples tested with the standard hose were shown to produce less variability.
4. Beam and cylindrical samples predicted field rutting performance approximately equally well.
5. Test temperature significantly affects measured rut depths in the APA. As test temperature increases, APA rut depths increase.
6. Test results when using the standard-diameter hose were collectively higher than were test results with the larger-diameter hose.
7. APA test results when using beam samples produced collectively higher rut depths than did cylindrical samples.

6.2 RECOMMENDATIONS

Based upon the test results and analysis, a tentative standard method of test in AASHTO format was developed and is recommended. This procedure is presented in Appendix B. Until a round-robin study is conducted using the automatic rut depth measuring system provided by the APA, the manual readings are tentatively recommended.

CHAPTER 7

VALIDATION OF PROPOSED APA TEST METHOD (PHASE III)

An experimental plan to validate the proposed test method developed in Phase II was conducted in Task 7. The plan included mixes with known field performance that were not used in the Phase II research. This chapter presents the experimental plan for Task 7 (Phase III), a description of the new mixes, test results, and analyses.

7.1 EXPERIMENTAL PLAN

The experimental plan for the validation phase included selecting 14 additional mixes of known field performance and testing these mixes using the APA test procedure recommended as a result of Phase II (see Appendix B). Two separate field pavement experiments were selected for this validation. The first was a full-scale field experiment on Interstate 80 (I-80) near Reno, Nevada. The Nevada DOT placed four different mixes on I-80: two were Superpave-designed mixes, and two were designed according to the standard Nevada DOT procedure (Hveem). These four mixtures were selected for the Phase III validation study. The second field experiment was the National Center for Asphalt Technology (NCAT) Test Track. Ten mixes were selected from this experiment.

Testing of the 14 Phase III mixes was carried out using both the cylindrical and beam specimens because the Phase II results did not differentiate one sample type from the other. All of the mixes were fabricated from original materials, proportioned to meet in-place properties, mixed in the laboratory, and subjected to short-term aging. Test temperatures for the mixes were selected as the high temperature of the standard performance grade. For all 14 mixes, this temperature was 64°C. For the NCAT Test Track mixes, replicated tests were performed using cylindrical and beams samples (two sets of six cylinders and two sets of three beams were tested) for selected mixes. The project panel recommended these replicated tests to evaluate the repeatability of the test methods. Only single replicate tests were conducted for the remaining mixes.

7.2 SELECTION OF MATERIALS

Based upon the approved Phase III experimental plan, 14 mixes of known rutting performance were included during the validation testing. The following sections describe these mixes.

7.2.1 Interstate 80 in Nevada (4 Test Sections)

Four test sections were placed on I-80 in Nevada near the WesTrack site in September 1998. The purpose of the experimental design for the four I-80 test sections was to compare the field performance of the Nevada DOT's (NDOT's) conventional Hveem mixture designed with an AC-20 P (polymer-modified AC-20) binder with Superpave mixtures placed at WesTrack. In addition, NDOT was interested in evaluating the effect of two factors—binder type and gradation—on the observed superior performance of its conventional mix over the Superpave mixes placed at WesTrack.

Figure 47 shows the layout of the test sections in the driving (westbound) lane of I-80 located at Milepost 63 east of Fernley, Nevada. Figure 48 shows the sampling and monitoring areas for these test sections. The condition survey of the test sections (including measurement of rut depths at 11 locations within the 500-ft-long monitoring section) was carried out at regular intervals by the University of Nevada at Reno. The last survey was conducted on April 4, 2001, after approximately 2.9 million ESALs.

Table 48 gives mix composition and average rut depths (measured on April 4, 2001) for the different test sections. The two asphalt binders used in the study were a Superpave PG 64-22, which was unmodified, and a viscosity-graded AC-20 that had been polymer modified (AC-20P). All four of the mixes had gradations that were 19.0-mm NMAS. As seen in Table 48, the two Superpave mixes were placed with almost identical gradations and the two NDOT (Hveem) mixes also had almost identical gradations (as intended). However, the Hveem-designed mixes were finer than the Superpave-designed mixes (Figure 49). On the 4.75-mm (No. 4) sieve, there was about a 10-percent difference in the two gradations. Also, the Superpave mixes had about 1 percent more asphalt binder than did the Hveem-designed mixes. Average rut depths were similar and relatively low for the two NDOT mixes as both had an average rut depth of 1.6 mm after approximately 2.9 million ESALs. The Superpave-designed mix using the PG 64-22 binder (unmodified) had the highest average rut depth at 15.1 mm; the Superpave-designed mix using the polymer-modified AC-20 had an average rut depth of 7.1 mm.

←

Direction of Traffic (westbound travel lane)

NDOT (AC-20P)	NDOT (PG 64-22)	Superpave (PG 64-22)	Superpave (AC-20P)

Figure 47. Layout of test sections on Interstate 80 near Milepost 63 near Fernley, Nevada.

Superpave (AC-20P) and Superpave (PG 64-22) Sections

175' Sampling	500' Monitoring	175' Sampling

NDOT (AC-20P) and NDOT (PG 64-22) Sections

100' Sampling	500' Monitoring	100' Sampling

Figure 48. Sampling and monitoring locations of Interstate 80 test sections.

7.2.2 NCAT Test Track (10 Test Sections)

The 1.7-mile oval NCAT Test Track was completed in August 2000; the loading began in September 2000. The test track had 26 test sections (each 60 m long) on the tangents, sponsored by nine states (Alabama, Florida, Georgia, Indiana, Mississippi, North Carolina, Oklahoma, South Carolina, and Tennessee) and FHWA. An additional 20 test sections were located on the two curves of the oval. Figure 50 shows the layout of the test sections. All of the test sections were constructed identically except for the top 100 mm (4 in.) that were used for the experimental mixes. Within the top 100 mm, test sections generally consisted of a 50-mm (2-in.) upper binder course and a 50-mm (2-in.) wearing course. All test sections were underlaid with 356 mm (14 in.) of HMA base course and 125 mm (5 in.) of permeable asphalt base. The following are specific mix attributes whose performance was compared on the test track:

- Coarse-graded, fine-graded, and through-restricted-zone gradation of Superpave mixes;
- Neat versus modified asphalt binder at optimum asphalt content as well as optimum +0.5 percent;

TABLE 48 Mix composition and rut data for I-80 test sections

Property	Test Section 1	2	3	4
Mix Type	**Superpave**	**Superpave**	**NDOT (Hveem)**	**NDOT (Hveem)**
Binder Type	**AC-20P**	**PG 64-22**	**PG 64-22**	**AC-20P**
Asphalt Content, %	**6.0**	**6.3**	**5.1**	**5.1**
Gradation, % passing				
25.0 mm	100	100	100	100
19.0 mm	99	99	93	95
12.5 mm	85	87	84	84
9.5 mm	73	75	75	74
4.75 mm	44	45	54	53
2.36 mm	28	29	40	40
1.18 mm	19	19	29	29
0.60 mm	14	14	21	22
0.30 mm	11	11	14	14
0.150 mm	8	9	9	9
0.075 mm	6.6	6.9	6.1	6.2
Average Rut Depth, mm, on April 4, 2001 (2.9 million ESALs)	7.1	15.1	1.6	1.6

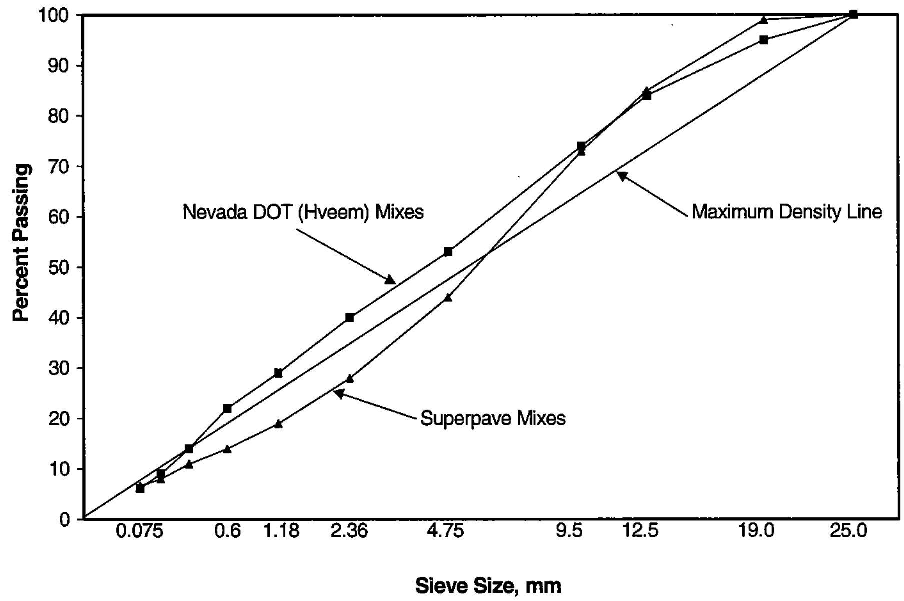


Figure 49. Gradations for Nevada Interstate 80 mixes.

- Stone matrix asphalt (SMA) versus Superpave mix using granite aggregate; and
- 12.5-mm versus 9.5-mm NMAS Superpave-designed mixes.

Mixes from 10 test sections of the track were selected for the validation research: N1, N3, N4, N11, N12, S4, S5, S8, S9, and S10. Selection of these sections was confirmed by the project panel to evaluate binder type and percentage (N1, N3, and N4); mix type (N11 and N12); aggregate type (S4, S5, and S8); and gradation shape (S9 and S10).

Table 49 provides the information on mix, aggregate, and binder type for the selected mixes. Nine of the ten mixes were designed according to the Superpave mix design sys-

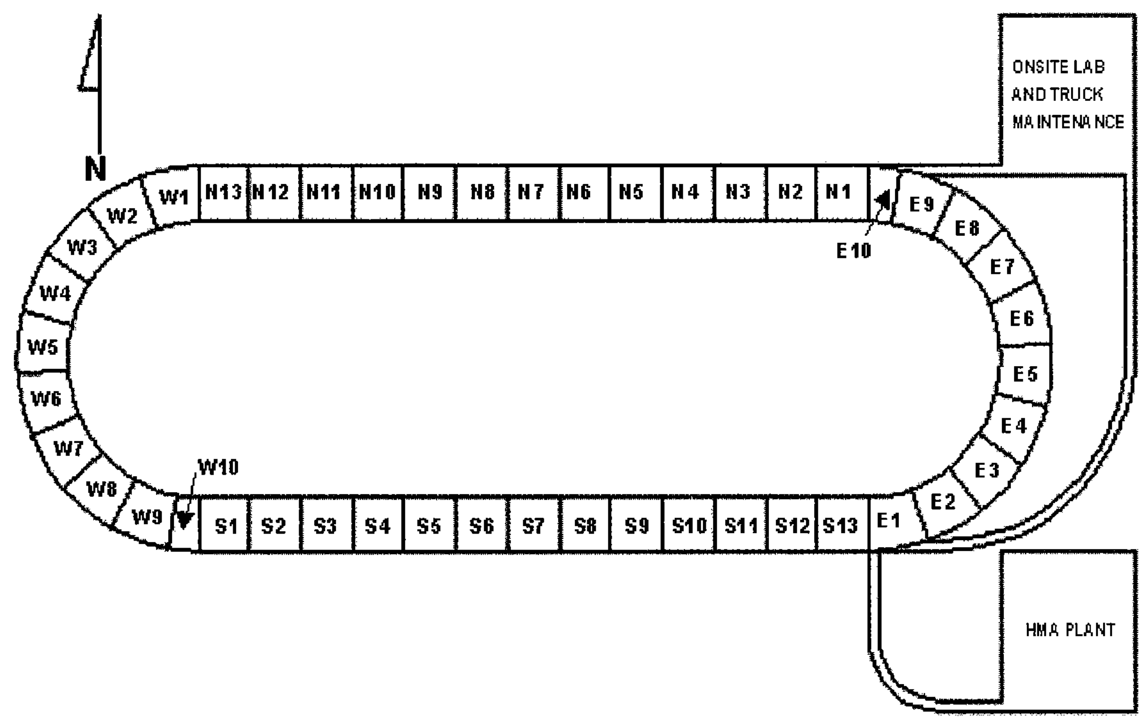


Figure 50. Layout of NCAT Test Track.

TABLE 49 In-place properties of selected NCAT Test Track sections

	SELECTED SECTIONS									
	N1	**N3**	**N4**	**N11**	**N12**	**S4**	**S5**	**S8**	**S9**	**S10**
Mix Type	Superpave	Superpave	Superpave	Superpave	SMA	Superpave	Superpave	Superpave	Superpave	Supepave
N_{design}	100	100	100	100	50 Blows	100	100	100	100	100
PG Binder	76-22 (SBS)	64-22	64-22	76-22 (SBS)	76-22 (SBS)	76-22 (SBS)	76-22 (SBS)	76-22 (SBS)	64-22	64-22
Aggregate Type	LMS/Slag	LMS/Slag	LMS/Slag	Granite	Granite	LMS/RAP	Cr. Gravel	Granite	Granite	Granite
Gradation Shape	ARZ	ARZ	ARZ	TRZ	SMA	ARZ	TRZ	BRZ	BRZ	ARZ
Sieve, mm	*% Passing*									
25.0	100	100	100	100	100	100	100	100	100	100
19.0	100	100	100	100	100	100	100	100	100	100
12.5	100	99	99	97	96	98	95	100	93	95
9.5	92	91	91	80	73	88	82	93	82	88
4.75	69	68	68	52	32	63	61	58	53	69
2.36	52	51	52	37	23	46	45	38	36	52
1.18	33	33	35	30	21	33	33	25	27	38
0.60	22	22	23	24	19	23	22	19	20	27
0.30	15	15	15	18	17	13	10	15	14	19
0.150	10	10	9	11	14	9	7	12	9	11
0.075	6.7	6.5	6.0	7.2	11.8	7.8	5.0	7.8	5.7	6.6
Asphalt Content	7.4	7.6	6.8	4.3	6.2	5.3	5.6	4.2	4.7	5.2
Average Density	95.1	94.1	93.4	93.1	94.6	94.3	94.9	91.8	93.4	93.7

NOTES: SMA = stone matrix asphalt; SBS = styrene-butadiene-styrene; LMS = limestone; RAP = reclaimed asphalt pavement; ARZ = above restricted zone; TRZ = through restricted zone; BRZ = below restricted zone.

tem. Of these nine, five mixes had gradations passing above the restricted zone (ARZ), two had gradations passing below the restricted zone (BRZ), and two had gradations passing through the restricted zone (TRZ). The tenth mixture was a stone matrix asphalt (SMA). Two different Superpave performance-graded asphalt binders were used with the ten mixes. Four mixes used an unmodified PG 64-22, and the remaining six mixes were produced with an SBS-modified PG 76-22. The PG 64-22 met the Superpave binder requirements beyond a high temperature of 67°C. Five of the ten mixes contained a granite aggregate, four contained limestone, and one mix was composed of a crushed siliceous gravel. Of the four mixes containing limestone (LMS), three were combined with a slag and the other included recycled asphalt pavement (RAP).

Table 49 shows that four of the mixes had 9.5-mm NMAS gradations and the remaining six mixes had 12.5-mm NMAS gradations. Figure 51 illustrates the gradations for the four 9.5-mm NMAS mixes. This figure shows that Sections N1, N3, and N4 had almost identical gradations. The primary differences among the mixes were binder type and content. The gradation for S8 was coarser than the mixes in N1, N3, and N4. NCAT Test Track mixes having a 12.5-mm NMAS are illustrated in Figure 52.

Table 49 also shows the average in-place density (expressed as a percentage of theoretical maximum density) of each section immediately after construction. All of the sections were constructed well: the lowest average density observed was 91.8 percent of theoretical maximum density for Section S8. Average rut depths for each of the sections after 8.9 million ESALs are presented in Table 50. This table shows that all of the test sections performed well with respect to rutting. The largest average rut depth for the ten sections was 6.0 mm (N3), which is less than half of the generally allowable rut depth of 12.5 mm (½ in.) during the pavement's service life. Seven of the ten test sections exhibited less than 2-mm rut depths, depths which are almost negligible.

7.3 TEST RESULTS AND ANALYSIS

The objective of Phase III was to validate the test method recommended during the Phase II testing. Table 51 presents the results of APA testing for the 14 mixes used in the validation. Within this table, the I-80 mixes from Nevada are provided in the first 4 rows, while the NCAT Test Track mixes are provided in the last 10 rows. Only a single replicate test was conducted for the Nevada mixes as there was insufficient material to conduct replicated tests. For the NCAT Test Track mixes, replicated tests were conducted on both the cylindrical and beam sample types for Sections N1, N3, and N4.

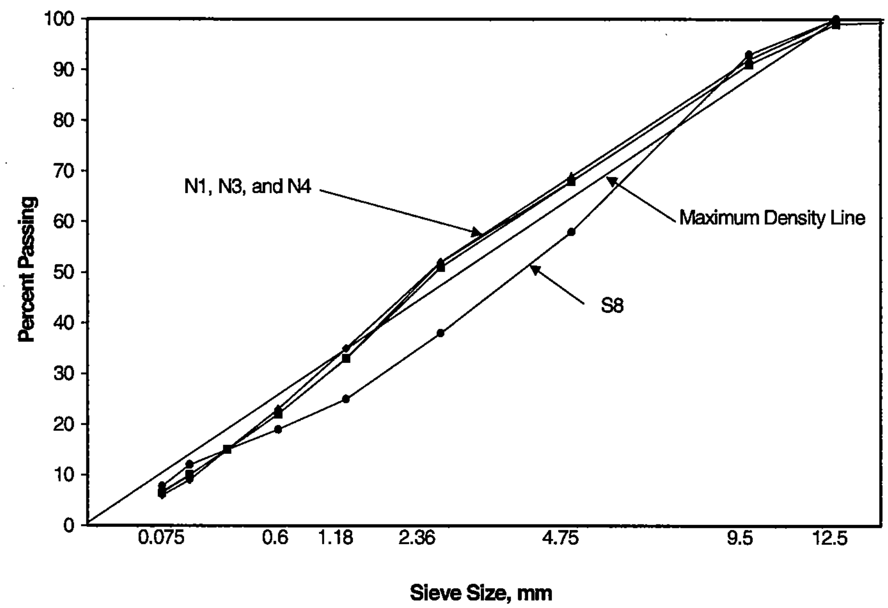


Figure 51. Gradations from NCAT Test Track having 9.5-mm NMAS gradations.

Replicated tests were also conducted with cylindrical samples for Section S4.

Figure 53 presents the relationships between field rut depths and APA rut depths for the Nevada I-80 mixes after 2.9 million ESALs for both the cylindrical and beam samples. Neither of the relationships was strong, having R^2 values of 0.60 and 0.45, respectively. The best fit lines shown in Figure 53 do follow the expected trend of increasing field rutting for increasing laboratory rut depths. However, the trend line for the cylindrical samples is steeper than is desired.

Figure 54 illustrates the relationship between field and laboratory rutting for the 10 NCAT Test Track sections. Only results for the first replicate testing are shown. Again, the R^2 values for the cylindrical and beam data sets were not strong at 0.23 and 0.002, respectively. The best fit line for the beam samples was basically flat, indicating no relationship between field and laboratory rutting (as would be expected with the very low R^2 value). Obviously, the APA did a poor job of predicting field performance for the test track mixes. All of the mixes placed on the test track could be considered premium,

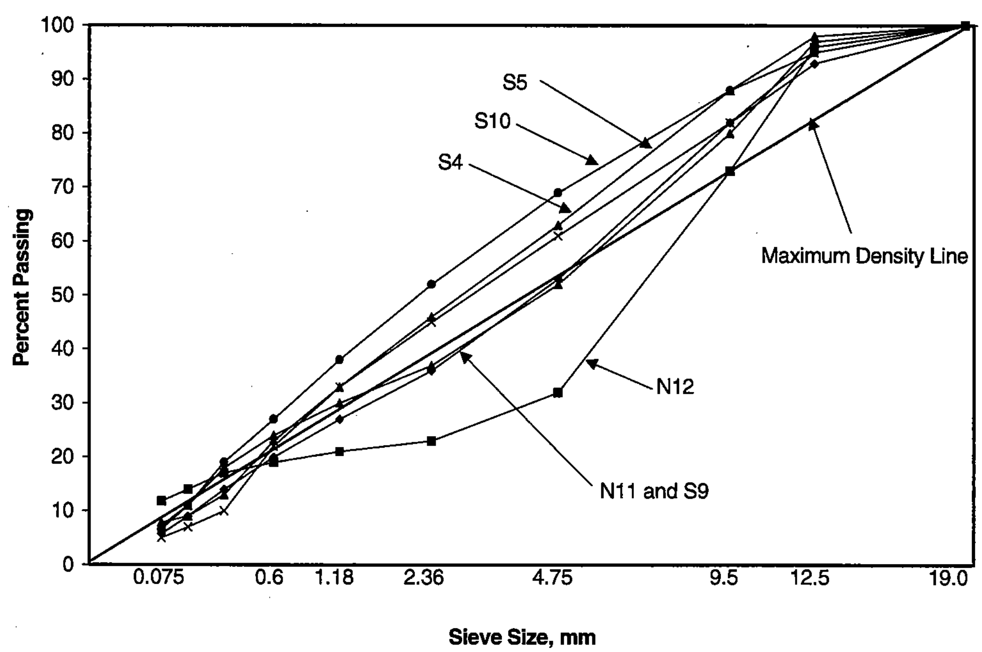


Figure 52. Gradations for NCAT Test Track having 12.5-mm NMAS gradations.

TABLE 50 Average rut depths for NCAT Test Track mixes after 8.9 million ESALs

Section	Average Rut Depth After 8.9 Million ESALs, mm
N1	2.2
N3	6.0
N4	4.0
N11	0.8
N12	1.5
S4	0.8
S5	0.8
S8	1.2
S9	1.0
S10	2.7

high-type mixes with the exception of Sections N2, N3, N5, N7, and N10. These sections were placed at binder contents approximately 0.5-percent higher than optimum. Of these, only N3 was included in the Phase III research. As shown in Table 50, Section N3 did have the highest rut depth in the field, at 6.0 mm. However, after 10 million ESALs, this magnitude of rutting was very small and may be nothing more than densification instead of rutting. According to APA test results on both beams and cylinders (Table 51), Section S5 should have rutted most; however, Section S5 rutted the least in the field. Therefore, if the rut performance exhibited by the sections is realistic of rutting that would take place under normal time and temperature effects, then the APA did not predict field performance.

The method for validating the proposed APA test method involved comparing the relationships between field and laboratory rutting developed during Phase II and Phase III research. However, there was a minor change to the Phase II data that first had to be accounted for. During Phase II, testing was conducted to 10,000 cycles and manual rut depths were obtained. In the proposed APA test method, 8,000 cycles were recommended. Therefore, testing during Phase III was conducted to 8,000 cycles.

In order to compare the Phase II and Phase III relationships between field and laboratory rutting, the manual rut depths obtained at 10,000 cycles during Phase II had to be adjusted to 8,000 cycles. In Chapter 5, the relationships between manual and automatic rut depth measurements at 8,000 and 10,000 cycles were discussed (see Figures 35 and 36). These relationships described in Chapter 5 were used to correct the Phase II relationships to 8,000-cycle manual rut depths.

The following sections compare the relationships between field and laboratory rutting for the Phase II and Phase III results. The hypothesis was that if the two relationships were similar, the relationship between field and laboratory rutting could be considered validated. This does not insinuate that the relationship was good or bad—only that the relationship between field and laboratory rutting was similar between

TABLE 51 APA test results for Phase III mixes

Mix	Cyl. Samples (Rep 1), mm	Cyl. Samples (Rep 2), mm	Beam Samples (Rep 1), mm	Beam Samples (Rep 2), mm
NDOT (AC-2OP)	4.36	—	6.46	—
NDOT (PG 64-22)	5.78	—	2.98	—
Superpave (AC-20P)	6.78	—	4.64	—
Superpave (PG64-22)	6.93	—	7.82	—
NCAT N1	1.40	3.60	3.43	3.15
NCAT N3	6.40	7.20	7.09	8.12
NCAT N4	8.30	7.43	4.42	5.27
NCAT N11	1.78	—	1.90	—
NCAT N12	3.40	—	3.84	—
NCAT S4	2.66	2.79	7.86	—
NCAT S5	7.97	—	13.29	—
NCAT S8	1.89	—	1.44	—
NCAT S9	4.85	—	4.98	—
NCAT S10	6.28	—	8.48	—

— Not tested.

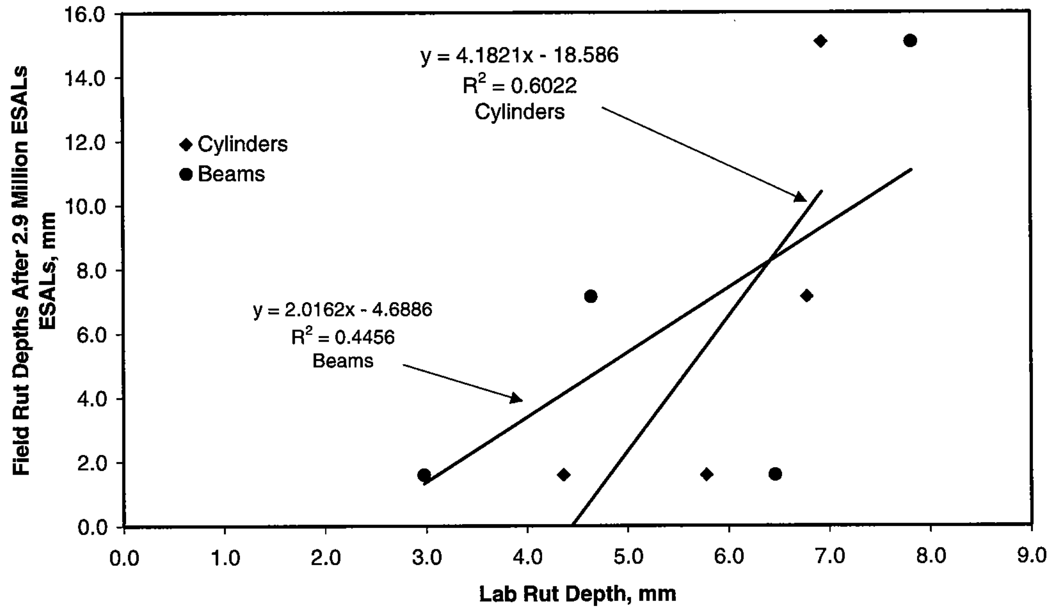


Figure 53. Relationship between field and lab rutting for Nevada I-80 mixes.

Phase II and Phase III. The strength of the relationship will define the usefulness of the APA in predicting the potential for rutting in the field.

7.3.1 Validation for Cylindrical Specimens

Comparison of the relationship between field and laboratory rutting during Phases II and III for cylindrical specimens is illustrated in Figure 55. The relationships shown in this figure are similar to the relationships shown for Phase II in that the field rut depths were normalized by the square root of the number of ESALs applied when the rut depth measurements were obtained. Figure 55 shows that the relationships between field and laboratory rutting are significantly different during Phases II and III, both in terms of intercepts and overall slopes. Collectively, the Phase III relationship had a flatter slope than did the Phase II relationship. Obviously, the Phase III data were influenced by very low field rut depths recorded on NCAT test sections.

One statistical method to compare whether the Phase III data improved the Phase II relationship between field and

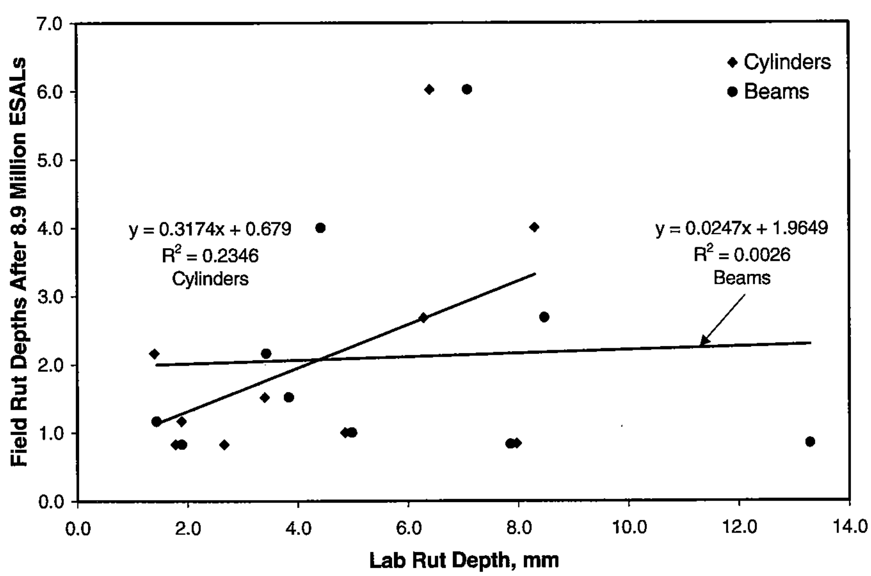


Figure 54. Relationship between field and lab rutting for NCAT Test Track mixes.

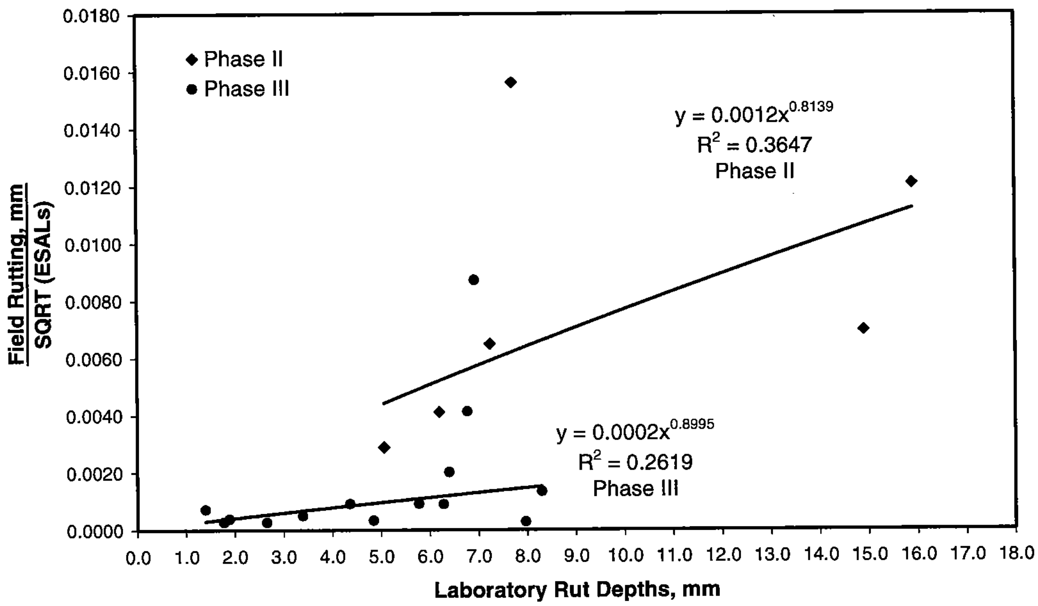


Figure 55. Comparison of Phases II and III relationship between field and lab rutting, cylinders.

laboratory rutting (therefore validating the Phase II conclusions) would be to input the Phase III data into the Phase II model (regression line) and evaluate the standard error of the estimate for all observations. The standard error of the estimate is the standard deviation of error (difference between observed value and predicted value) about a model. If the standard error of the estimate decreases with the inclusion of the Phase III data, the Phase II conclusions can be considered validated. This analysis is presented in Table 52. Based upon the standard error of the estimate for the Phase II data alone (0.00455) and the standard error of the estimate for the combined Phase II and III data (0.00338), the inclusion of the

TABLE 52 Comparison of standard errors of the estimate for cylindrical samples

Phase	Section	APA Rut Depth, mm	Rutting Rate[a]	Predicted Rutting Rate	Error	Standard Error of the Estimate
	C16	5.10	0.002885	0.004497	–0.00161	
	C20	15.90	0.012076	0.011401	0.00068	
	C21	14.90	0.006950	0.010815	–0.00386	
	S15	6.20	0.004113	0.005297	–0.00118	
	S19	7.20	0.006482	0.006013	0.00047	
II	S24	7.70	0.015610	0.006333	0.00928	0.00455[b]
	NVAC-20P	4.36	0.000917	0.003978	–0.00306	
	NVPG64-22	5.78	0.000917	0.005004	–0.00409	
	SPAC-20P	6.78	0.004128	0.005698	–0.00157	
	SPPG64-22	6.93	0.008715	0.005800	0.00292	
	S4	2.66	0.000278	0.002661	–0.00238	
	S5	7.97	0.000281	0.006499	–0.00622	
	S8	1.89	0.000391	0.002015	–0.00162	
	S9	4.85	0.000335	0.004338	–0.00400	
	S10	6.28	0.000896	0.005354	–0.00446	
	N1	1.40	0.000723	0.001578	–0.00086	
	N3	6.40	0.002014	0.005437	–0.00342	
	N4	8.30	0.001338	0.006718	–0.00538	
	N11	1.78	0.000278	0.001919	–0.00164	
III	N12	3.40	0.000508	0.003249	–0.00274	0.00338[c]

[a] Field rut depth divided by square root of ESALs.
[b] Standard error of the estimate including only Phase II data.
[c] Standard error of the estimate including both Phase II and Phase III data using Phase II regression.

Phase III data improved the relationship between field and laboratory rutting. Based upon these standard errors of the estimate, the Phase II relationship between field and laboratory rutting was validated statistically with the Phase III data.

Figure 56 illustrates the relationship between field and laboratory rutting when both the Phase II and Phase III data are combined. The relationship between field and laboratory rutting was significant (a *P*-value of less than 0.05) although the R^2 was only 0.47. Also, the R^2 value increased (from 0.36 to 0.47) for the combined data set compared with the Phase II data alone.

7.3.2 Validation for Beam Specimens

Comparison of the relationship between field and laboratory rutting during Phases II and III for beam specimens is illustrated in Figure 57. In this figure, it appears that the relationships between field and laboratory rutting were significantly different (in terms of general slopes) during Phases II and III. Similar to the cylindrical samples (see Figure 55), the Phase III relationship had a flatter slope than did the Phase II relationship.

Table 53 presents the standard errors of the estimate for the Phase II data alone and combined Phase II and III data. Based on the standard error of the estimate for the Phase II data alone (0.00415) and the standard error of the estimate for the combined Phase II and III data (0.00334), the inclusion of the Phase III data did improve the relationship between field and laboratory rutting. Based upon these standard errors of the estimate, the Phase II relationship between field and laboratory rutting was validated statistically with the Phase III data.

Figure 58 illustrates the relationship between field and laboratory rutting when both the Phase II and Phase III beam data are combined. The relationship between field and laboratory rutting was significant (a *P*-value of less than 0.05) although the R^2 was only 0.34. Unlike cylinders, the R^2 value decreased (from 0.50 to 0.34) for the combined data set compared with the Phase II data alone.

7.3.3 Validation Analysis Considering Potential Outliers

The WesTrack Section 24 mix was identified as a possible outlier during Phase II. Since none of the 10 NCAT Test Track sections developed any significant rutting in the field, all 10 could potentially be considered outliers. Extensive analysis as described in preceding Sections 7.3.1 and 7.3.2 were repeated excluding these data of potential outliers. However, the analysis results were generally not different from those including all sections.

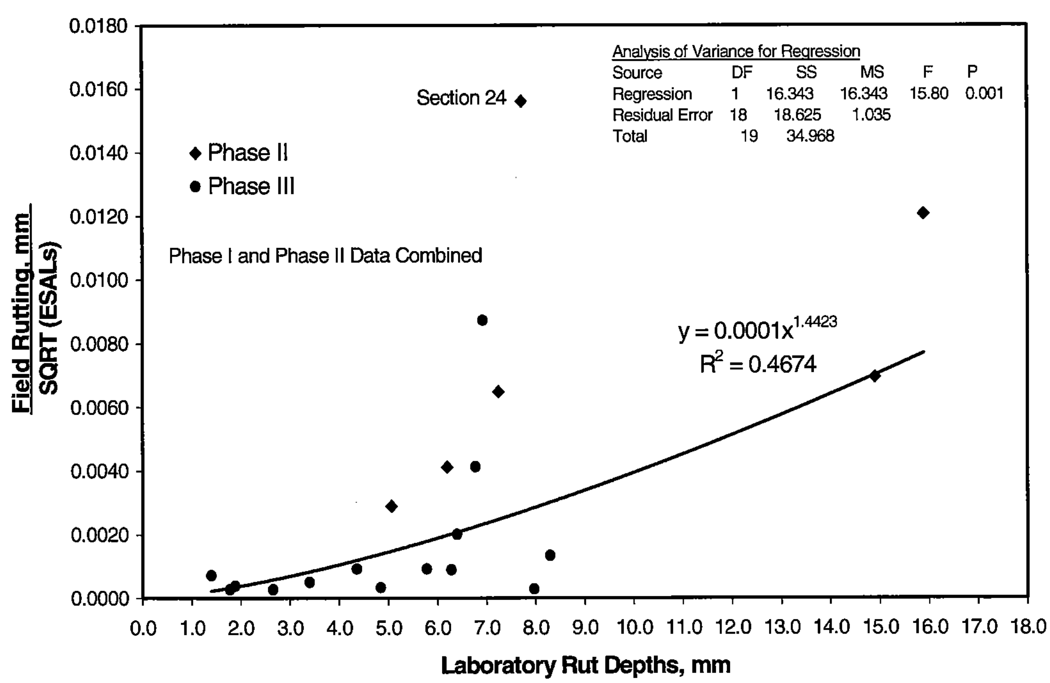


Figure 56. Combined Phase II and Phase III relationship between field and laboratory rutting (4PGSC).

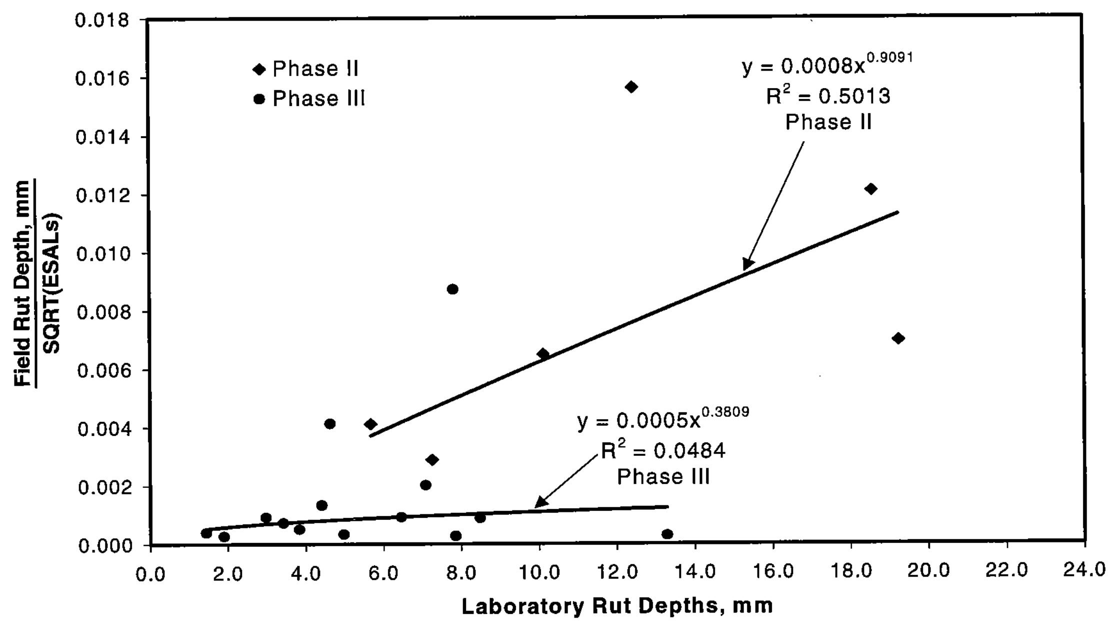


Figure 57. Comparison of Phase II and Phase III relationships between field and lab rutting (5PGSB).

7.3.4 Validation Conclusions

It was concluded in Phase II that the APA can be used to determine the rutting potential of HMA mixes in the field. Although statistical analyses in Sections 7.3.1 and 7.3.2 have generally indicated validation of this Phase II conclusion in Phase III, it is evident that the absolute relationship between the APA rut depths and the field rut depth varies from project to project depending on each project's geographical location and traffic characteristics. Therefore, it is not possible to predict the field rut depths from the APA rut depths on a specific project using relationships from other projects in a different environment. It is also possible that the relationship is influenced by the mix type.

TABLE 53 Comparison of standard errors of the estimate for beam samples

Phase	Section	APA Rut Depth, mm	Rutting Rate[a]	Predicted Rutting Rate	Error	Standard Error of the Estimate
	C16	7.25	0.002885	0.004846	–0.001961	
	C20	18.57	0.012076	0.011390	0.000686	
	C21	19.24	0.006950	0.011764	–0.004814	
	S15	5.68	0.004113	0.003878	0.000235	
	S19	10.12	0.006482	0.006560	–0.000077	
II	S24	12.44	0.015610	0.007916	0.007694	0.004153[b]
	NVAC-20P	6.46	0.000917	0.004362	–0.003444	
	NVPG64-22	2.98	0.000917	0.002159	–0.001241	
	SPAC-20P	4.64	0.004128	0.003229	0.000900	
	SPPG64-22	7.82	0.008715	0.005189	0.003526	
	S4	7.86	0.000278	0.005213	–0.004936	
	S5	13.29	0.000281	0.008404	–0.008123	
	S8	1.44	0.000391	0.001114	–0.000723	
	S9	4.98	0.000335	0.003443	–0.003109	
	S10	8.48	0.000896	0.005586	–0.004689	
	N1	3.43	0.000723	0.002453	–0.001731	
	N3	7.09	0.002014	0.004747	–0.002733	
	N4	4.42	0.001338	0.003089	–0.001751	
	N11	1.90	0.000278	0.001434	–0.001156	
III	N12	3.84	0.000508	0.002718	–0.002210	0.003338[c]

[a] Field rut depth divided by square root of ESALs.
[b] Standard error of the estimate including only Phase II data.
[c] Standard error of the estimate including both Phase II and Phase III data using Phase II regression.

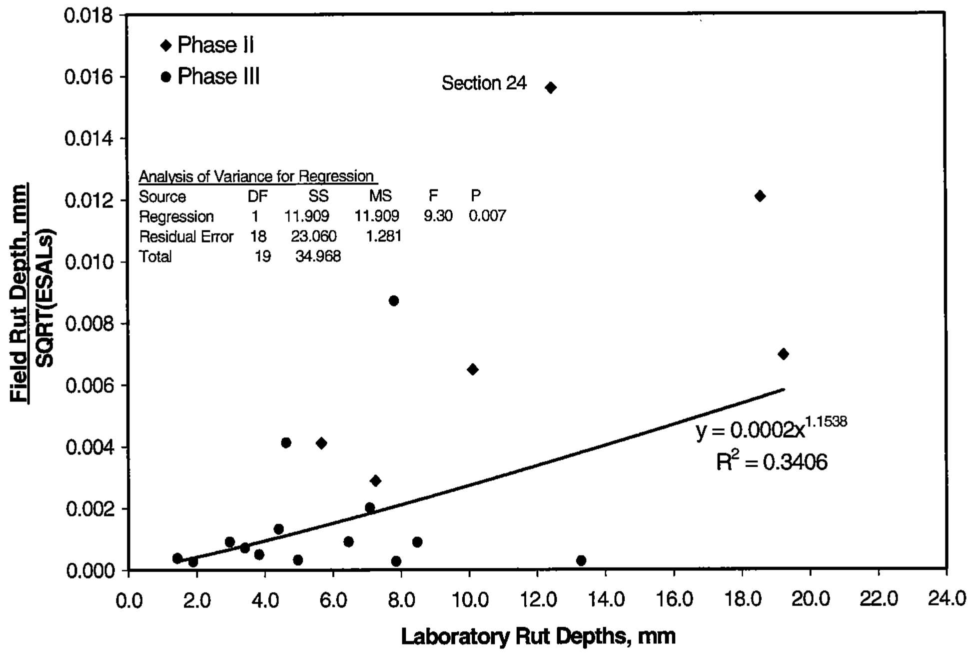


Figure 58. Combined Phase II and Phase III relationship between field and laboratory rutting (5PGSB).

TABLE 54 Single-factor ANOVA cylindrical samples

Source	Degrees of Freedom	Sum of Squares	Mean Squares
Total	7	48.869	
Mix	3	45.742	15.257
Error	4	3.127	0.782

TABLE 55 Single-factor ANOVA for beam samples

Source	Degrees of Freedom	Sum of Squares	Mean Squares
Total	5	61.325	
Mix	2	60.395	30.197
Error	3	0.931	0.310

7.4 ESTIMATE OF REPEATABILITY

Using the data provided in the previous sections, the validation of the Phase II relationships between field and laboratory rutting were evaluated. The strengths of the relationships for both cylinders and beams were not strong, but they were significant. The strength of the relationships were reasonable considering the influences of climatic conditions, varying traffic effects, and varying materials at each of the field projects used within this study. One remaining question is that of repeatability. To investigate repeatability, the replicated tests shown in Table 51 were used. A single-factor ANOVA was conducted for both sample types with the mix being the single factor. The ANOVA was selected because the mean squares error produced by the ANOVA is a pooled estimate of repeatability variance. By taking the square root of the mean squares error term, a pooled estimate of repeatability standard deviation can be obtained. Tables 54 and 55 present the ANOVA results for the cylindrical and beam sample types, respectively.

In Tables 54 and 55, the beam samples had a lower pooled estimate of repeatability variance (0.31 for beams and 0.78 for cylinders). This also indicates a smaller pooled estimate of repeatability standard deviation (0.56 for beams and 0.88 for cylinders). Therefore, the replicated tests showed more repeatable results for beam sample types.

CHAPTER 8

PROPOSED RESEARCH PLAN TO EVALUATE APPLICABILITY OF APA FOR FIELD QC/QA OPERATIONS

8.1 INTRODUCTION

Currently the Superpave mix design system does not include any tests to evaluate the rutting susceptibility of HMA. The purpose of NCHRP Project 9-17 was to evaluate the APA as a suitable proof test for determining the rutting susceptibility of designed HMA mixes. According to the completed study, the APA is an acceptable method for proof testing laboratory-designed mixes.

The usefulness of any mix design proof test is extended if it can be used in QC/QA applications. An owner agency must know whether a produced HMA will perform satisfactorily on the roadway. Ideally, this type of information should be available within 24 h or less of a mix being produced. Therefore, it is important that the APA be evaluated for its usefulness as a field QC/QA test method. Before the APA can be used as a QC/QA test method, the following issues should be researched further.

8.1.1 Sample Size

NCHRP Project 9-17 recommended cylindrical samples compacted to 4 ± 0.5-percent air voids or beam samples compacted to 5 ± 0.5-percent air voids. For both sample types, sample height was set at 75 ± 3 mm. Compacting samples to a specified air void content is generally an iterative process. Samples must be compacted and allowed to cool, and then the bulk-specific gravity must be determined in order to obtain air void contents. If the samples do not meet the desired air void–content range, then another set of samples must be compacted. This process can take a number of iterative steps in order to obtain enough samples for testing.

In order to simplify the compaction of cylindrical samples and to speed up the testing of QC/QA samples, a possible solution would be to simply compact samples to the design number of gyrations. A potential problem with this compaction method is that samples are 115 ± 5 mm in height. This is different from the 75 mm recommended in NCHRP Project 9-17. Research should be conducted to compare rut depths of samples compacted to 4-percent air voids with samples compacted to the design number of gyrations. Also included in this work would be a comparison between the two sample heights.

8.1.2 Hose Pressure and Wheel Load

NCHRP Project 9-17 recommended using a wheel load and hose pressure equal to 533 N (120 lb) and 830 kPa (120 psi), respectively. Historically, APA testing has been conducted at 445 N (100 lb) and 690 kPa (100 psi), respectively. There is concern by some that a large proportion of APAs do not currently have an adequate air pressure supply to maintain an 830-kPa hose pressure. The differences in rut depths for testing at 533 N/830 kPa and 445 N/690 kPa were not investigated during NCHRP Project 9-17. An investigation should be conducted to compare rut depths between the two loading conditions to determine whether there is a high correlation. Because of the recommendations of a higher than historical test temperature (the high temperature of the standard performance grade for a project location), it is possible that the 445 N/690 kPa loading may be sufficient.

8.1.3 Effect of Reheating Prior to Compaction

In quality assurance applications, plant mix is commonly cooled down, taken to a central laboratory, reheated, and then compacted for testing. Samples that have to be reheated in this manner may have further aging, leading to a stiffer mix. An investigation should be conducted to compare rut depths of samples that are "hot compacted" with samples that are reheated. This should be conducted on plant-mixed HMA.

8.1.4 Effect of Asphalt Absorption

The production of HMA can be a relatively quick process. Concerns have recently been raised that when an HMA sample is tested immediately after production, there is insufficient time for aggregates to absorb the asphalt binder similar to what occurs during the short-term oven aging procedure during mix design. In essence, this results in higher effective asphalt binder contents than in mix that is transported and placed on the roadway. The transport and placement allows time for more asphalt binder to be absorbed.

This problem may affect results of APA testing during QC/QA. If a mixture contains an aggregate with a relatively high absorption value, the mix may act like it is over-asphalted if compacted immediately after production. This, in turn, would lead to high rut depths; however, these high rut depths may not be indicative of the mix that was placed on the roadway. Some states have addressed this problem by requiring a 1-h oven aging on plant-produced mix. An experiment should be conducted to determine whether plant-produced mix needs oven aging prior to compaction and, if so, at what level of absorption does oven aging become necessary. This experiment should contain aggregates with a wide range of water absorption values.

8.1.5 Comparison of Laboratory and Plant-Produced Rut Depths

One issue that must be addressed in a QC/QA study of the APA is a comparison of rut depths for laboratory- and plant-produced mixtures. The Superpave mix design system does a reasonably good job of simulating the condition and characteristics of HMA after field production. However, mix that goes through an HMA production plant is still different than laboratory-prepared mix. A comparison in rut depths for identical mixes (gradation and asphalt content) between field- and laboratory-produced mix is needed. This comparison should provide insight into development of critical APA rut depths during QC/QA operations.

CHAPTER 9

RECOMMENDED PRACTICE FOR ESTABLISHING MAXIMUM SPECIFIED RUT DEPTH FOR APA

9.1 OBJECTIVE

The objective of this recommended practice is to give highway agencies a method of calibrating APA rut depth criteria for local climate and traffic.

9.2 SCOPE OF RECOMMENDED PRACTICE

There are two prevailing methods of calibrating laboratory permanent deformation tests to field rutting. The first entails testing mixes during production and then following the performance of these mixes over time. This method is time consuming, but provides a more accurate field calibration. The second method entails identifying existing pavements with a wide range of rutting performance. Samples of the pavement are cut from the roadway, and the aggregates are extracted. An asphalt binder similar to the original binder is then combined with the extracted aggregate and aged in the laboratory, and performance testing is conducted. Results of the testing are then compared with performance in the field. The following sections describe these two calibration procedures.

9.3 TESTING OF PLANT-PRODUCED MIX

Identify HMA projects to be constructed that fall within the four primary traffic categories (as shown below). At each of the projects, compact samples of plant-produced mix to meet the sample requirements of the APA draft standard procedure. Cylinders, beams, or both can be investigated, depending upon the agency. At least four pavements should be tested for each traffic category. In order to evaluate repeatability, enough samples of the same mix should be compacted to conduct replicate tests (one replicate equals six cylindrical samples or three beams).

Traffic Category	20-year Design ESALs	N_{design} Gyrations
Very high	>30 million	125
High	3–30 million	100
Medium	0.3–3 million	75
Low	<0.3 million	50

9.3.1 Evaluation of Test Data and Development of Critical Rut Depths

For all traffic categories of asphalt pavements sampled, tabulate the data as shown in the following table. Use engineering judgment in reviewing all the data in the table, and establish a minimum APA rut depth specification requirement for each traffic category to ensure good rutting performance. The specification must take into account the repeatability and reproducibility of the APA test, if available.

Traffic Category*	Rutting Performance	Average Field Rut Depth (mm)	Average APA Rut Depth (mm)
Very high	good		
	fair		
	poor		
High	good		
	fair		
	poor		
Medium	good		
	fair		
	poor		
Low	good		
	fair		
	poor		

* Categories should recognize traffic speed, climatic conditions, and structural influences.

9.4 TESTING OF EXISTING ASPHALT PAVEMENTS

Identify at least three asphalt pavements (or overlays) that have been in service from 3 to 5 years in the following four 20-year design traffic categories.

Traffic Category	20-year Design ESALs	N_{design} Gyrations
Very high	>30 million	125
High	3–30 million	100
Medium	0.3–3 million	75
Low	<0.3 million	50

The pavements in each traffic category should be selected to provide the following rutting performance in the field after 3 to 5 years in service: good (less than 5 mm of rut depth); fair (5–10 mm of rut depth); and poor (over 10 mm of rut depth). Therefore, a minimum of 12 asphalt pavements should be sampled. The number in some or all traffic categories can be increased to improve confidence in specified acceptable rut depth criteria for APA.

Obtain HMA mix from each pavement by coring or sawing, which should be done within 600 mm (2 ft) of the pavement edge (outside wheel path) to represent as-placed HMA as much as possible. Sampling from wheel tracks is not desirable because of potential degradation of the HMA under traffic. If cores are obtained, the cores should be at least 150 mm in diameter to minimize inclusion of aggregate particles cut by the coring operation. Sampling should be done on a level stretch of the highway and within the region where the field rut depth was recorded. Enough material samples should be obtained to produce the following:

- Six SGC specimens 150 mm diameter × 75 mm height *or* three beam specimens 300 mm × 125 mm × 75 mm,
- Three loose mixture samples (1500 gram each) to determine the theoretical maximum density (TMD), and
- Three loose mixture samples (2500 gram each) to determine the asphalt content.

Obtain 40-percent more material than needed above to account for wastage, retests, or both.

9.4.1 Analysis of In-Place Mix

Conduct three ignition tests on the HMA sample obtained from each asphalt pavement to obtain the average asphalt content and average gradation of the in-place mix.

Please note that if desired, bulk-specific gravity of the core or sawed samples and TMD of the in-place mix can be measured to determine the in-place air voids for information only.

9.4.2 Preparation of Test Samples

Conduct solvent extraction on the sampled, in-place mixture to extract aggregate for preparing fresh mixture using virgin asphalt binder. Obtain a virgin asphalt binder with the same performance grade as is used on the project sampled. If a modified binder was used on the project, obtain a similarly modified binder of the same performance grade.

Mix the extracted aggregate and the virgin performance-grade binder to obtain the average in-place asphalt content in the mix. Subject the prepared mix to short-term aging at the desired compaction temperature suited for the performance grade being used. Conduct three replicate tests to determine the average TMD of the aged mix, which will be used to control the air void content in the compacted specimens.

Compact six SGC samples to obtain 4 ± 0.5-percent air void content in the samples. (Agencies that prefer beams should compact three beams at 5 ± 0.5-percent air void content.) Where possible, replicate tests should be conducted.

9.4.3 Testing by APA

The six SGC specimens or three beam specimens should be tested to determine the average rut depth after 8,000 loading cycles. Testing should be done at the high temperature of the performance grade recommended for the project location regardless of bumping. For example, a polymer-modified PG 76-22 or PG 70-22 may have been used on a project that required a PG 64-22 corresponding to local climatic conditions. In that case, APA testing should still be conducted at 64°C.

9.4.4 Evaluation of Test Data and Development of Specifications

For all traffic categories of asphalt pavements sampled, tabulate the data as shown below. Use engineering judgment in reviewing all the data in the table, and establish a minimum APA rut depth specification requirement for each traffic category to ensure good rutting performance. The specification must take into account the repeatability and reproducibility of the APA test, if available.

Traffic Category*	Rutting Performance	Average Field Rut Depth (mm)	Average APA Rut Depth (mm)
Very high	good		
	fair		
	poor		
High	good		
	fair		
	poor		
Medium	good		
	fair		
	poor		
Low	good		
	fair		
	poor		

* Categories should recognize traffic speed, climatic conditions, and structural influences.

CHAPTER 10

STUDY CONCLUSIONS AND RECOMMENDATIONS

10.1 CONCLUSIONS

The conclusions provided below are based upon the research results obtained during research in Phases II and III:

- Cylindrical samples compacted to 4-percent air voids and beam samples compacted to 5-percent air voids resulted in APA laboratory test results that were more closely related to field rutting performance than did cylindrical and beam samples compacted to 7-percent air voids (Phase II).
- Samples tested in the APA at a test temperature corresponding to the high temperature of the standard performance grade for a project location better predicted field rutting performance than did samples tested at 6°C higher than the high temperature of the standard performance grade (Phase II).
- Samples tested with both the standard- and large-diameter hoses predicted field rutting performance about equally well. However, samples tested with the standard hose produced less variability (Phase II).
- Beam and cylindrical samples predicted field rutting performance about equally well (Phase II).
- Test temperature significantly affects measured rut depths in the APA. As test temperature increases, APA rut depths increase (Phase II).
- APA-measured rut depths were collectively higher with the standard-diameter hose than with the larger-diameter hose (Phase II).
- APA-measured rut depths were collectively higher with beam samples than with cylindrical samples (Phase II).
- Using the preceding conclusions, in this study an improved test protocol was developed for the APA in order to better identify rut-prone HMA mixtures (Phase II).
- Laboratory rut depths measured by the APA had good correlations on an individual project basis with the field rut depths in the case of FHWA ALF, WesTrack, MnRoad, and I-80 (Nevada) projects. However, the APA-measured rut depths had a poor correlation with field rut depths in the case of 10 test sections on the NCAT Test Track, which did not develop any significant rutting after 2 years of loading (Phases II and III).
- Based upon limited data, the APA compared well with other performance tests with respect to predicting the potential for rutting in the field (Phase II).
- It is generally not possible to predict field rut depths from APA rut depths on a specific project using relationships developed on other projects with different geographical locations and traffic (Phase III).

10.2 RECOMMENDATIONS

Using the test results and analysis, a tentative standard method of test in AASHTO format was developed and is recommended. This procedure is presented in Appendix B. A round-robin study is needed to better evaluate repeatability and reproducibility of the proposed test method.

REFERENCES

1. Brown, E.R., and S.A. Cross. "A National Study of Rutting in Hot Mix Asphalt (HMA) Pavements." *Proceedings of the Association of Asphalt Paving Technologists*, Vol. 61; 1992; pp. 535–582.
2. Lai, J.S. "Development of a Simplified Test Method to Predict Rutting Characteristics of Asphalt Mixes." Final Report, Research Project No. 8503, Georgia DOT, July 1986.
3. Lai, J.S. "Evaluation of the Effect of Gradation of Aggregate on Rutting Characteristics of Asphalt Mixes." Project No. 8706, Georgia DOT, August 1988.
4. Lai, J.S. "Evaluation of the Effect of Gradation of Aggregate on Rutting Characteristics of Asphalt Mixes." Project No. 8609, Georgia DOT, December 1986.
5. West, R.C., G.C. Page, and K.H. Murphy. "Evaluation of the Loaded Wheel Tester." *Research Report FL/DOT/SMO/91-391,* December 1991.
6. Lai, J.S. "Results of Round-Robin Test Program to Evaluate Rutting of Asphalt Mixes Using Loaded Wheel Tester." *Transportation Research Record No. 1417,* Transportation Research Board of the National Academies, Washington, DC, 1993; pp. 127–134.
7. Lai, J.S., and H. Shami. "Development of Rolling Compaction Machine for Preparation of Asphalt Beam Samples." *Transportation Research Record No. 1492,* Transportation Research Board of the National Academies, Washington, DC, 1995; pp. 18–25.
8. Shami, H.I., J.S. Lai, J.A. D'Angelo, and T.P. Harmon. "Development of Temperature Effect Model for Predicting Rutting of Asphalt Mixtures Using Georgia Loaded Wheel Tester." *Transportation Research Record No. 1590,* Transportation Research Board of the National Academies, Washington, DC, 1997; pp. 17–22.
9. "Rutting Susceptibility of Bituminous Mixtures by the Georgia Loaded Wheel Tester." *Report No. RDT98-001*, Missouri Department of Transportation, May 1, 1998.
10. Collins, R., H. Shami, and J.S. Lai. "Use of Georgia Loaded Wheel Tester to Evaluate Rutting of Asphalt Samples Prepared by Superpave Gyratory Compactor." *Transportation Research Record No. 1545*, Transportation Research Board of the National Academies, Washington, DC, 1996; pp. 161–168.
11. Choubane, B., G.C. Page, and J.A. Musselman. "Investigation of the Asphalt Pavement Analyzer for Predicting Pavement Rutting." *Research Report FL/DOT/SMO/98-427,* 1998.
12. Kandhal, P.S., and R.B. Mallick. "Evaluation of Asphalt Pavement Analyzer for HMA Mix Design." *NCAT Report 99-4,* National Center for Asphalt Technology, June 1999.
13. West, R.C. "A Ruggedness Study of the Asphalt Pavement Analyzer Rutting Test." Memorandum to the Asphalt Pavement Analyzer User Group and New APA Owners, May 14, 1999.
14. Prowell, B.D. "Development of Rutting Criteria for the Asphalt Pavement Analyzer." Paper submitted for consideration for presentation and publication at the 1999 International Conference on Accelerated Pavement Testing; Paper No. CS6-4; 1999.
15. Huber, G.A., J.C. Jones, P.E. Messersmith, and N.M. Jackson. "Contribution of Fine Aggregate Angularity and Particle Shape to Superpave Mixture Performance." *Transportation Research Record No. 1609,* Transportation Research Board of the National Academies, Washington, DC, 1998; pp. 28–35.
16. Stuart, K.D., and R.P. Izzo. "Correlation of Superpave $G^*/\sin\delta$ with Rutting Susceptibility from Laboratory Mixture Tests." *Transportation Research Record No. 1492*, Transportation Research Board of the National Academies, Washington, DC, 1995; pp.176–183.
17. Stuart, K.D., and W.S. Mogawer. "Effect of Compaction Method on Rutting Susceptibility Measured by Wheel-Tracking Devices." Paper presented at the 76th Annual Meeting of the Transportation Research Board, Washington, DC, January 12–16, 1997.
18. Williams, C.R., and B.D. Prowell. "Comparison of Laboratory Wheel-Tracking Test Results with WesTrack Performance." *Transportation Research Record No. 1681,* Transportation Research Board of the National Academies, Washington, DC, 1999; pp. 121–128.
19. Wu, P. "Contact Pressure Measures of GLWT and Astec Asphalt Pavement Analyzer." Interdepartmental Correspondence, Office of Materials and Research, Georgia DOT, July 12, 1996.
20. Lai, J.S. "Development of a Laboratory Rutting Resistance Testing Method for Asphalt Mixes." Project No. 8717, Georgia DOT, August 1989.
21. Aurilio, V., T. Pellinen, E.T. Harrigan, and K. Stuart. "Preliminary Validation of Superpave Using the Accelerated Loading Facility (ALF)." *Proceedings of the Canadian Technical Asphalt Association, Forty-First Annual Conference,* Edmonton, Alberta, November 1996.
22. Stuart, K.D., W.S. Mogawer, and P. Romero. "Validation of Asphalt Binder and Mixture Tests That Measure Rutting Susceptibility Using Accelerated Loading Facility." *Report No. FHWA-RD-99-204, Interim Report*, October 1993–October 1999.
23. Cooley, L.A., Jr., and P.S. Kandhal. "Evaluation of Density Gradients in Loaded Wheel Tester Samples." *Journal of Testing and Evaluation*, Vol. 28, No. 6; November 2000; pp. 427–484.

GLOSSARY

AASHTO: American Association of States Highway and Transportation Officials
ALF: Accelerated Loading Facility
ANOVA: analysis of variance
APA: Asphalt Pavement Analyzer
ARZ: above restricted zone (gradation)
AVC: Asphalt Vibratory Compactor

BBR: bending beam rheometer
BRZ: below restricted zone (gradation)

COV: coefficient of variation

DOT: department of transportation
DMRT: Duncan's Multiple Range Test
DSR: dynamic shear rheometer

$E^*/\sin\delta$: dynamic modulus
ESALs: equivalent single axle loads

FHWA: Federal Highway Administration
$F_{(N)}$: flow number as measured by a repeated load test
***F*-statistic:** test of significance for variances
$F_{(T)}$: flow time as measured by triaxial creep test

$G^*/\sin\delta$: the viscous component of the binder shear stiffness, as measured by AASHTO TP5 and used as a specification parameter in AASHTO MP1
GDOT: Georgia DOT
GIT: Georgia Institute of Technology
GLWT: Georgia Loaded Wheel Tester

HMA: hot mix asphalt
HWTD: Hamburg Wheel Tracking Device

LCPC: Laboratoire Central des Ponts et Chaussées (France's central lab for highways)
LMS: limestone
LWT: Loaded wheel tester

MnRoad: Minnesota Road Research Project
***m*-value:** the rate of change with time of the creep stiffness, *S*, as measured by AASHTO TP1 and used as a specification parameter in AASHTO MP 1

NCAT: National Center for Asphalt Technology
NCHRP: National Cooperative Highway Research Program
N_{design}: design number of gyrations
NDOT: Nevada DOT
NMAS: nominal maximum aggregate size

PAV: pressure aging vessel as described in AASHTO PP1
PCC: Portland cement concrete
PG: performance grade

QC/QA: quality control/quality assurance

R^2: correlation coefficient
RAP: recycled asphalt pavement; asphalt paving material milled from existing bituminous pavement, consisting of aggregate and asphalt binder
RTFO: rolling thin film oven

SBS: styrene-butadiene-styrene polymer
SGC: Superpave Gyratory Compactor
SMA: stone matrix asphalt
SPT: simple performance test

TFHRC: Turner-Fairbank Highway Research Center
TMD: theoretical maximum density
TRB: Transportation Research Board
TRZ: through restricted zone (gradation)

APPENDIX A
LITERATURE REVIEW

Task 1 of the project reviewed the literature for information on the state of practice for the Asphalt Pavement Analyzer (APA) and other loaded wheel testers (LWTs). This review was to address each LWT's suitability for predicting the rutting potential of hot mix asphalt (HMA) during laboratory mix design and quality control/quality assurance (QC/QA) testing. Critical test parameters, limitations, material sensitivities, and boundary conditions of the various LWTs were to be identified. Additionally, the review was to be used to identify areas for the APA requiring further evaluation and/or standardization to verify or improve its ability to predict rutting potential.

The literature review was conducted to specifically answer the following questions:

- What are the key test parameters, limitations, material sensitivities, and boundary conditions used by various LWTs?
- What are the conclusions and recommendations of researchers who have evaluated various LWTs, specifically the suitability of LWTs in predicting rutting?
- What areas need further evaluation and standardization to verify or improve the APA's ability to predict rutting?

The predominant LWTs found in the literature were the French LWT, Hamburg LWT, and APA (or Georgia LWT). However, some references were found that included other types of LWTs. Additionally, some publications compared different LWTs during research. The first section of this appendix deals with literature on the French Rutting Tester. The second section reviews information concerning the Hamburg LWT or modified versions (e.g., the Couch or Superfos Construction Rut Tester). The third section describes literature on the APA, including references on the Georgia LWT. The fourth section describes other LWTs for which only one or two references are available, as well as references that may have information that is relevant to NCHRP Project 9-17 but do not specifically include LWT testing. The fifth section provides information on research studies that compared two or more LWTs.

A.1 FRENCH LOADED WHEEL TESTER

The Laboratoire Central des Ponts et Chaussées (LCPC) wheel tracker (also known as French Rutting Tester) has been used in France for over 15 years to evaluate the rutting characteristics of HMA. In recent years, the French Rutting Tester (FRT) has been used in the United States, most notably by the state of Colorado and at the FHWA's Turner Fairbank Highway Research Center, to evaluate the rutting and stripping potential of HMA.

The FRT is capable of testing two HMA slabs. Slab dimensions are typically 180 mm wide, 500 mm long, and 100 to 200 mm thick. Testing consists of loading a pneumatic tire inflated to 600 kPa with 5,000 N. The tire then loads a sample at a rate of one cycle per second; each cycle includes a forward and backward stroke. A "zero" rut depth is generally defined by loading a specimen at an ambient temperature for 1,000 cycles. Slab specimens are then heated to between 35 and 60°C for 12 h prior to testing. After temperature conditioning, slabs are loaded for 30,000 cycles, which takes about 9 h.

Deformation depths are typically taken after 100; 300; 1,000; 3,000; 10,000; and 30,000 cycles. Deformation is defined as the average of a series of 15 measurements consisting of 3 measurements taken across the width of the specimen at five locations along the length of the beam. Rut criteria are typically the average deformation expressed as a percentage of the original slab thickness.

Nevelt and Thanfold (*1*) described work that was accomplished to evaluate different road structures and asphalt mixes using the FRT. Road structure indicates the testing of multilayer (binder and wearing courses) specimens. Testing conducted in the FRT was similar to the standard method discussed above except specimens were preheated for 24 h prior to testing.

Nevelt and Thanfold presented the results of five testing programs. For the first program, a motorway from Austria was used. Cores were extracted from the motorway and recombined into beams for testing. The selected motorway was one of the most heavily trafficked in Austria and had rut depths of 3 mm after 16 months. Results of FRT testing indicated that rutting was 6 percent after 30,000 cycles. This was less than the typically recommended 10-percent maximum for highly trafficked roadways. Therefore, the authors concluded that for the Austrian motorway, the FRT accurately predicted that the HMA structure would resist deformation.

The second test program compared two different roadways' structures. Tests were conducted on both the wearing and binder courses individually for each project and also on multilayer specimens. Results indicated that the multilayer testing compared very well with the individual layer testing for both projects.

The third test program tested base course mixes at different test temperatures. Identical mixes were tested at both 50 and 60°C. One of the base course mixes used a polymer-modified asphalt binder while the other did not. From the

FRT results, the authors concluded that the polymer-modified base course mix was less susceptible to temperature. The percent rut depths for the base course mix containing the polymer additive were almost identical at both test temperatures (4.9 and 5.1 percent at 50 and 60°C, respectively), while there was a difference of 5.3 percent in rut depths for the unmodified base course mix at the two test temperatures.

The fourth test program compared two stone matrix asphalt (SMA) mixes with different mortar (filler, asphalt binder, and additives) contents. The primary difference in mortars was the percentage of filler. Results indicated that the FRT could differentiate among mortars.

The final test program compared four different wearing course mixes. Two of these mixes were SMAs, and two were dense-graded. Results indicated that the FRT could differentiate among the four wearing course mixes.

Because of this review, the primary experimental variable that was included within NCHRP Project 9-17 was test temperature. Nevelt and Thanfold showed that identical mixes tested at 50 and 60°C provided differences in laboratory-measured rutting.

Aschenbrener (*2*) presented the results of a study to correlate the FRT to actual pavement performance. The FRT used in this study tested confined slabs with a length of 500 mm and width of 180 mm (20 and 7.2 in., respectively). Thicknesses of 20 to 100 mm could be tested. Loading conditions were identical to the standard operating characteristics presented previously. Prior to testing, each slab was aged at room temperature for as long as 7 days prior to obtaining the "zero" rut depth.

Aschenbrener presented some stress conditions found in FRT test slabs. Using the pavement analysis program CHEVPC, the average compressive stress throughout the thickness of a 100-mm slab was 410 kPa (60 psi). In a 50-mm slab, the average compressive stress was 550 kPa (80 psi).

Aschenbrener described three different approaches that could be used to compare FRT results with actual field performance. The first entailed testing newly designed mixes in the FRT and then placing that mix on a project. The field project would then be monitored for rutting over time, and the field results would be compared with laboratory results. The second approach was to obtain field cores and/or slabs from projects of known performance for testing. In the third approach, original raw materials from projects of known performance could be obtained, re-blended in the laboratory, and tested. Results reported by Aschenbrener were obtained using the second approach.

Selection of the 33 pavement sections by Aschenbrener was based upon rutting performance, temperature, and traffic. Within Colorado, three Superpave high-temperature environments existed. High-temperature environments were defined by the highest monthly mean maximum temperatures.

Traffic was characterized by the equivalent daily 18-kip load applications (EDLAs) instead of equivalent single axle loads (ESALs). In Colorado's experience, a pavement can be performing satisfactorily and then rut almost immediately within 1 month in a hot summer. Therefore, rut depths do not increase linearly with cumulative ESALs. Actual traffic loading at the time the rut depth increases significantly would be the most desirable value, but Aschenbrener indicated that the information was not available. Additionally, since rut depths do not generally show significant increases after this immediate rutting, total ESALs are not appropriate. Aschenbrener therefore decided that EDLAs provided a better comparison of relative traffic loadings for each level of highway analyzed.

Field rut depths were obtained from Colorado's network-level pavement management system. Sites with high and low levels of rutting were selected for each of the traffic and temperature classes.

From each pavement selected, three slabs were obtained for testing in the FRT. One slab was generally tested at 50°C (122°F), and a second slab was tested at 60°C (140°F). The third slab was tested at either 40 or 45°C (104 or 113°F) for pavements from low-temperature sites or 55°C (131°F) for pavements from moderate- to high-temperature sites.

Results indicated that the 60°C test temperature showed promise as a "go, no go" criterion for pavements located in high-temperature locations. However, pavements placed in moderate-temperature locations were significantly affected by the actual test temperature in the FRT. By reducing the test temperature by 10°C, six of the pavements that had good field performance went from failing to passing while no pavements with poor performance went from failing to passing. An interesting finding by the author was that correlations between low-temperature pavements and field performance were highly variable. However, when the elevation of the pavement site was taken into account, correlations were much stronger.

Table A-1 presents the coefficient of correlation (R^2) values for predicting actual rutting depths with the FRT for different combinations of test temperature and traffic level. Aschenbrener (2) investigated both the slope of the rutting curve (B) and the log of the test cycles to failure (C/1,000) obtained from the FRT data. For each test temperature, the addition of traffic levels to the model greatly increased the correlation between laboratory and field results. Results in Table A-1 do not include testing of pavements from low-temperature sites because these were not tested at 60°C.

Based upon the research results, Aschenbrener suggested new "go, no go" specifications for the FRT. He indicated that test temperatures should match the actual pavement temperatures. Therefore, he recommended that 50, 55, and 60°C be used as test temperatures for the three temperature regions within Colorado. Interestingly, using these test temperatures, none of the pavements studied that had poor field performance would have been placed in the field. Only 4 of 14 well-performing pavements would have been eliminated.

This informative paper provided several potential variables for inclusion within NCHRP Project 9-17:

TABLE A-1 Coefficients of correlation (R^2) for predicting actual rutting depths with FRT results (*2*)

	Slope (*B*)	Log (*C*/1,000)
60°C Test Temperature		
All Traffic	0.45	0.47
> 400 EDLA	0.67	0.74
< 400 EDLA	0.65	0.68
> 250 EDLA	0.61	0.69
< 250 EDLA	0.60	0.72
50°C Test Temperature		
All Traffic	0.37	0.44
> 400 EDLA	0.52	0.75
< 400 EDLA	0.84	0.78
> 250 EDLA	0.47	0.61
< 250 EDLA	0.80	0.71
60° or 50°C Test Temperature		
All Traffic	0.49	0.35
> 400 EDLA	0.87	0.70
< 400 EDLA	0.68	0.48
> 250 EDLA	0.67	0.61
< 250 EDLA	0.72	0.38
60° or 55°C Test Temperature		
All Traffic	0.45	0.33
> 400 EDLA	0.78	0.76
< 400 EDLA	0.70	0.56
> 250 EDLA	0.60	0.63
< 250 EDLA	0.72	0.50

- What is the test temperature?
- What criteria should be used within an APA critical rut depth specification? This report provides traffic and temperature as potential variables. Additionally, Aschenbrener suggests that for lower-traffic roadways, the number of cycles during the test could be reduced.
- How should testing samples be obtained for comparing laboratory results with field performance?

A.2 HAMBURG WHEEL TRACKING DEVICE

The Hamburg Wheel Tracking Device (HWTD) was developed in Hamburg, Germany. It is used as a specification requirement for some of the most heavily trafficked roadways in Germany to measure rutting and stripping. The HWTD test slabs are 260 mm wide, 320 mm long, and 40 mm thick. The HWTD also has the ability to test field cores. Testing is accomplished underwater at temperatures ranging from 25 to 70°C, with 50°C (122°F) being the most commonly used temperature. A 47-mm-wide steel wheel loaded by 705 N (158 lb) is used to test samples. Specimens are loaded for 20,000 passes or until 20 mm (0.8 in.) of deformation occurs. The travel speed of the wheel is approximately 1.2 km/h.

Results from the HWTD consist of a rut depth, creep slope, stripping slope, and stripping inflection point. The creep slope is the inverse of the deformation rate in the linear region of the deformation curve after preliminary consolidation and prior to stripping. The stripping slope is the inverse of the deformation rate after stripping has occurred. Stripping inflection point is defined as the number of passes corresponding to the intersection of the creep and stripping slopes.

A slight variation to the HWTD is the Superfos Construction Rut Tester (SCRT). The SCRT uses slab specimens with similar dimensions as the HWTD. The primary difference between the SCRT and HWTD is the loading mechanism. The SCRT applies an 82.6-kg (180-lb) vertical load to a solid rubber wheel (instead of to a steel wheel) with a diameter of 194 mm and a width of 46 mm. This loading configuration results in an applied contact pressure of approximately 950 kPa (140 psi); it is applied at a rate of approximately 2 km/h. Testing is typically conducted at 55°C (131°F). Results from the

SCRT are identical to the HWTD and include rut depth, creep slope, stripping slope, and stripping inflection point.

Messersmith, Jones, and Wells (*3*) described the evaluation of polymer-modified asphalt binders in Alabama with the SCRT. Slab specimens were tested in the SCRT using a hard rubber wheel that has an approximate contact pressure of 950 kPa (140 psi) with a contact area of 8.25 cm^2 (1.28 $in.^2$). Slab specimens were compacted to approximately 6-percent air voids and tested at 60°C (140°F). Results of the laboratory rutting experiment indicated that the SCRT was able to differentiate between the different polymer additives.

This paper identified several potential variables for possible inclusion within NCHRP Project 9-17:

- The contact area of the hard rubber wheel with the slab specimen: although the contact area was not varied in this study, different contact areas could produce different results; and
- Contact pressure, compactive effort, percent air voids, and temperature.

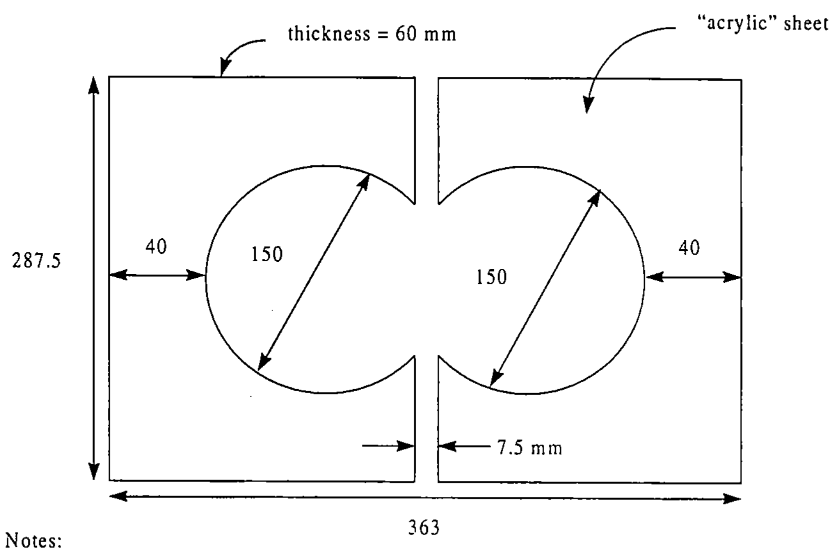
Izzo and Tahmoressi (*4*) presented the results of a study using the HWTD to (1) evaluate the repeatability of the HWTD, (2) compare slab and cylindrical specimens, and (3) evaluate temperature and anti-stripping additives. This study primarily investigated moisture susceptibility in HMA, so it is not totally applicable to NCHRP Project 9-17. However, the authors suggest a new test specimen configuration for cylindrical specimens that could be considered in NCHRP Project 9-17.

The authors compared traditional HWTD slab specimens to Superpave Gyratory Compactor (SGC) specimens, configured as shown in Figure A-1. The cylindrical specimens are sawed to approximately 25 mm so that they can be "butt" up against each other. Results of the comparison study indicated that the magnitude of the stripping inflection point was different between the traditional slabs and "new" cylindrical specimens; however, the two sample types did rank the different mixtures in a similar manner. This new sample configuration could be a variable that needs consideration within the APA when cylindrical samples are used.

Hall and Williams (*5*) presented the results of a study to evaluate the effects of specimen type, compaction type, and sample configuration on LWT results. The LWT used in this study was a Hamburg-type piece of equipment built by the Department of Civil Engineering at the University of Arkansas and was known as the Evaluator of Rutting and Stripping in Asphalt (ERSA) machine.

ERSA testing was conducted on samples submerged in water at 50°C (122°F). A 47-mm (1.9-in.) wide steel wheel loaded specimens with 705 N (160 lb) for 20,000 cycles or a 20-mm rut depth, whichever occurred first. Hall and Williams' test matrix included cylindrical and beam samples cut from newly compacted HMA pavements and SGC samples compacted from companion loose mixture.

Field slabs with 600 mm × 600 mm (24 × 24 in.) dimensions were cut from pavements to produce four 150 mm × 300 mm (6 × 12 in.) beams. Plaster was then used to produce specimens 380 mm long × 300 mm wide × 175 mm thick (15 × 12 × 7 in.) dimensions for testing. Field cores (150 mm) were also cut from the pavements for testing. Gyratory specimens (150 mm) were compacted from loose plant mix to between 6- and 8-percent air voids.



Notes:
1. not to scale
2. dimensions in millimeters

Figure A-1. Top view of Superpave Gyratory Compactor specimen configuration for the HWTD (4).

Several different analyses were conducted by the authors. First, the effect of specimen type was investigated using the field-cut beams and cores. Only field-cut specimens were included in the analysis so that compaction method was not a variable. Results of this analysis showed that for all surface mixes tested, no statistical difference was found in rut depths between the two sample types. Hall and Williams did stipulate that these results represent specimens compacted in the same manner to the same air void content and taken from the same location in the pavement.

The next analysis was to evaluate the effect of compaction method. For this analysis, the authors lumped the field-cut beams and cores into one category called field compaction and compared results with gyratory-compacted samples. Results indicated that rut depths for the field- and gyratory-compacted specimens were significantly different. Field-compacted specimens exhibited larger rut depths than did the laboratory-compacted specimens.

The final analysis was conducted to determine whether the effect of sawing a flat surface on cylindrical specimens (similar to Figure A-1) so that the specimens could be butted against each other provided significantly different results when compared with the normal separated testing configuration. This analysis was only conducted on gyratory specimens. Results indicated that no differences occurred between the standard sample configuration (separated) and the sample configuration with sawed surfaces that butt against each other. However, the authors did note that significant differences did occur in the stripping susceptibility of the two configurations.

The following variable was identified for possible inclusion within NCHRP Project 9-17:

- When comparing laboratory rut depths with field rut depths, should APA testing be conducted on field-compacted specimens or laboratory-compacted specimens?

A.3 APA

The initial version of the APA (then called the Georgia LWT) was developed in 1985 through a partnership between the Georgia DOT (GDOT) and the Georgia Institute of Technology. The Georgia LWT, originally designed to test slurry seals, was modified to perform efficient, effective, and routine laboratory testing and field production quality control of HMA. The Georgia LWT has been used by GDOT and approximately 10 other state DOTs for determining the expected performance of their laboratory-designed HMA mixtures with regards to rutting resistance.

The Georgia LWT is capable of testing beam or cylindrical specimens. Beams are typically 125 mm wide, 300 mm long, and 75 mm thick. Cylindrical specimens are typically 150 mm in diameter and 75 mm thick. Loads are applied to test specimens through a wheel onto a pneumatic hose, which rests on top of the test specimen. A wheel load of 445 N (100 lb) and a hose pressure of 690 kPa (100 psi) are common. Test temperatures normally range from 40 to 60°C. Testing is generally conducted to 8,000 cycles.

The APA is the second-generation commercial version of the Georgia LWT. The APA has additional features to allow the machine to evaluate not only the rutting potential of mixes, but also their moisture susceptibility and fatigue cracking under service conditions. The APA follows the same rut-testing procedure used for the Georgia LWT.

Lai (*6*) documented the development of the LWT that later became known as the Georgia LWT (GLWT). The study was conducted to (1) develop a simple apparatus to predict rut-susceptible mixes, (2) evaluate rut-prediction capability of the developed apparatus, and (3) compare results from the developed apparatus with results from creep and repeated load triaxial tests.

The test program consisted of developing the apparatus, preparing samples for testing, testing the samples using a developed test procedure, and conducting comparison testing with creep and repeated load triaxial tests. An existing LWT already located in the GDOT Materials Testing Laboratory and used to test slurry seals was the basis for the GLWT. To enhance the rut-predicting capabilities of this apparatus, several modifications were made.

First, a 76-mm (3-in.) linear hose and an aluminum wheel combination replaced a 203-mm (8-in.) diameter aluminum wheel with a 25-mm (1-in.) diameter hose wrapped around it. This was done because the original aluminum wheel tended to skid near the ends of the samples. The linear hose also seemed to show less wear and could be easily replaced. The second modification was to adapt the slurry seal testing apparatus to accommodate 76 mm × 76 mm × 380 mm (3 in. × 3 in. × 15 in.) beams of HMA. The next task was to devise a method for measuring rut depths. This was accomplished by developing a template that contained seven slots and fit over the sample mold. A micrometer was then used to obtain measurements prior to and then after testing. Differences between these measurements were defined as rut depth. The third modification was to strengthen the mounting table to provide a stable base during testing.

In order to evaluate the ability of the apparatus to predict rut-susceptible mixes, Lai used four mixes from Georgia with known field rut performance. Three of these four mixes had shown a tendency to rut in the field. Both cylindrical and beam samples were used in the evaluation. A "loaded foot" kneading compactor was used to compact both specimen types. Cylindrical samples were compacted by "spooning" hot mixture into the mold as the loaded foot assembly was compacting the mixture. Beam samples were compacted in three equal lifts. A sliding rack was needed for the beams because the kneading compactor was stationary. As each layer was placed into the beam mold, the mold was moved as the loaded foot compacted. After all three layers were compacted, a loading plate was placed onto the top of the beam and statically loaded until the appropriate height of 76 mm was achieved.

The test procedure used by Lai entailed preheating samples to 35°C (95°F) within an air-tight room for 24 h. This room was built to maintain 35°C, and the LWT was located within this room. Prior to testing, the rut profile of the beam sample was determined with the template and micrometer.

Lai performed an experiment using two linear hose pressures and three applied wheel loads. Pressures used were 689 and 517 kPa (100 and 75 psi), and wheel loads were 445, 334, and 222 N (100, 75, and 50 lb). Wheel loads were applied by placing steel plates into a weight holding box located above the wheel. To characterize rutting, the author measured rut depths at 0; 40; 100; 400; 1,000; and 4,000 cycles. One cycle equaled a forward and backward pass of the loaded wheel over the linear hose. During the discussion of the development of the equipment, Lai indicated that the loaded wheel was operating at approximately 44 cycles per minute.

Results of the testing showed that all samples had similar symmetrical rutting profiles transversely across the sample tops. This indicated that rutting was only taking place underneath the linear hose. However, the longitudinal profile showed uneven rutting. Rut depths on the beam nearest the pivot of the reciprocating arm with the wheel were highest. This was attributed to the downward shoving action of the arm when pushing the wheel forward. Lai therefore averaged rut depths along the middle portion of all specimens.

As would be expected, rut depths increased as the number of cycles increased. Also, the higher wheel load and higher hose pressure generally resulted in more severe rutting (see Figure A-2). At the lowest wheel load and pressure, the effect of mix type was almost indistinguishable.

When comparisons were made between the results of the developed LWT and results of creep and repeated load triaxial testing, Lai gave the opinion that the LWT seemed to predict rutting more in line with what was found in the field.

Based upon the research performed in this study, Lai concluded that the developed LWT was relatively simple to perform and that the results obtained had the potential for being able to assess the rutting characteristics of asphalt mixes. Lai further recommended that several modifications be made to the developed LWT:

1. Modify the reciprocating arm to alleviate the excessive rutting at the end of the beam samples;
2. Replace the linear hose with a pneumatic tire (bicycle type);
3. Develop a better device to measure rut depths; and
4. Include an environmental chamber to maintain specimen temperature.

This report was of significant historical importance because it was the beginning of what is now known as the GLWT and, subsequently, as the APA. Important factors or variables for the possible inclusion within NCHRP Project 9-17 include the following:

- What combination of wheel load and hose pressure should be used to best predict the rutting potential of HMA mixtures?
- What method of compaction should be used for samples?
- What specimen geometry should be used—beam or cylindrical?

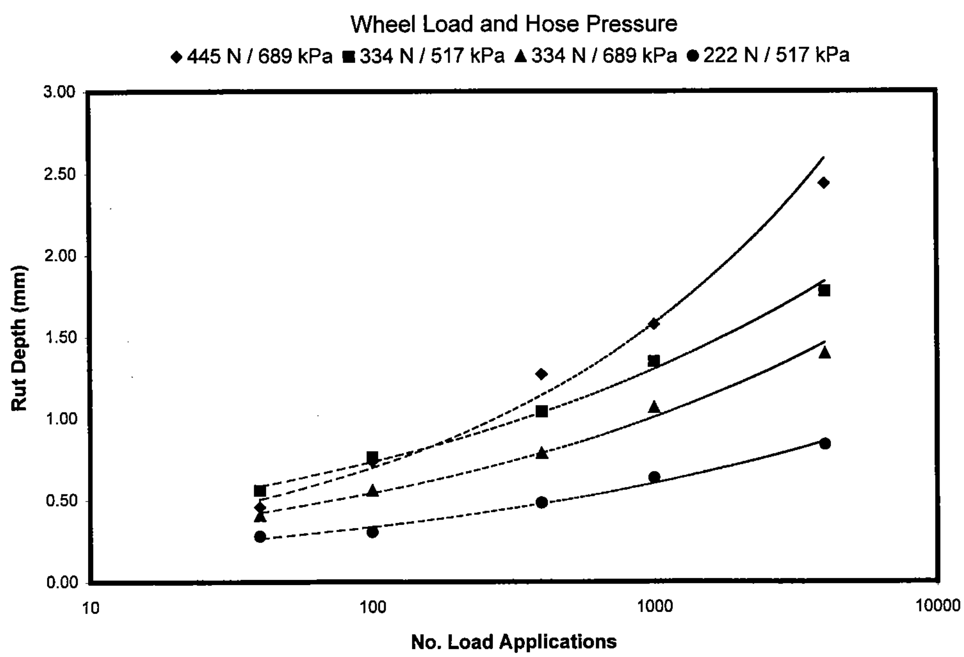


Figure A-2. Lai's test results for Mix A (6).

- What comparison tests should be included? (Lai used creep and repeated load triaxial tests.)

Lai (*7*) documented a study conducted to evaluate GDOT "B" type mixes and six different modified mixes. These six modified mixes used different mineral fillers common in Georgia. Also, one mix used an asphalt binder modified with 6-percent polymer.

Preparation of beam specimens was similar to the previous paper (*6*) in that a loaded-foot kneading compactor was used. Testing conditions included a test temperature of 35°C (95°F), wheel load of 445 N (100 lb), hose pressure of 689 kPa (100 psi), and load frequency of 22 cycles per minute—a substantial change from that used previously by Lai (*6*). Rut depth measurements were obtained at 0; 200; 500; 1,000; and 2,000 cycles. However, some mixtures were left in the GLWT for as many as 10,000 cycles.

Results of this study provided information to GDOT as to which HMA mixes (or modifiers) showed less potential for rutting. In the recommendations, Lai did mention several deficiencies pertaining to the GLWT test method used for this study. The author mentioned sample size, sample confinement, test temperature, and method of sample preparation as needing more research. The GDOT B mix from Lai's July 1996 paper (*7*) used larger aggregates than did the mixes used in the initial research on the GLWT (*6*). Also, because of the larger aggregates, sample preparation was more difficult. Beams used in this study were 75 mm × 75 mm × 380 mm, and the author indicated that because of the larger aggregate sizes, the coarse aggregate particles in the beam developed an interlocking action in combination with the confinement of the beam molds preventing plastic deformation and lateral shoving of the mix.

Several variables identified by Lai in his recommendations were considered in NCHRP Project 9-17:

- What is the proper sample size? Lai indicated that for mixes containing larger aggregates, the beam sample size should be increased. If this is correct, can the beam dimensions be decreased for mixes with smaller aggregates?
- What is the proper test temperature? Lai mentioned that test temperature should be investigated further. He did not elaborate what test temperatures should be investigated, but did point out the importance of test temperature.
- What boundary conditions exist in the GLWT (or APA)? Similar to proper sample size, what dimensions are needed to best simulate mix behavior in the field?
- What method of sample preparation should be used? For this study, samples were conditioned for 24 h at test temperature.

Lai (*8*) presented the results of a study to evaluate the effect of varying both the maximum nominal aggregate size and the fine aggregate portion of a gradation on rutting potential. Maximum aggregate sizes of 19.0 to 37.5 mm (3/4 to 1 1/2 in.) were investigated. Percentages of aggregate passing the 2.36-mm (No. 8) *sieve* ranged from 22 to 38 percent. Preparation of the beam samples was similar to procedures used in Lai's earlier work (*6, 7*). Testing conditions included a test temperature of 40.6°C (105°F), a wheel load of 445 N (100 lb), a hose pressure of 689 kPa (100 psi), and a frequency of loading of 22 cycles per minute. Rut depths were measured after 0; 200; 500; 1,000; 2,000; 3,000; 4,000; 5,000; 6,000; 7,000; and 8,000 cycles.

Results of the study provided information on which combinations of maximum aggregate size and fine aggregate proportion reduced rutting potential. Interestingly, mixes using an aggregate source with a relatively higher percentage of elongated particles showed the highest potential for rutting.

During testing, Lai noted some problems in manually measuring rut depths. Instances occurred in which coarse aggregates were immediately underlying the template's groove. This led to some rut depth measurements that were lower than expected; the author disregarded these readings and averaged the remaining rut depths.

Several potential variables were identified from this report for the possible inclusion within NCHRP Project 9-17:

- Lai increased the test temperature over previous studies (*6, 7*). The question then arises as to which test temperature is most correct.
- Based upon the problems with measuring rut depths with the manual template, the best method of measuring rut depths could be evaluated.
- Within the recommendations of the report, Lai indicated that mixtures containing larger aggregate sizes should be evaluated further. The main concern indicated by Lai was the test sample size in relation to the aggregate size.

Lai (*9*) conducted an investigation for GDOT to further modify and improve the GLWT and its sample preparation method. Also, a standardized test procedure was developed. Primary improvements or modifications to the GLWT noted in this report included the addition of an environmental chamber to make the equipment self-contained and the development of a preheating box for preconditioning samples. Additional improvements included the addition of a load cycle counter, improved safety measures, and a rut profile–measuring template.

An interesting experiment described was the evaluation of linear hoses with different "stiffness." In Lai's previous studies (*6, 7, 8*), a flexible-type rubber hose was used. The author surmised that a stiffer hose could affect test results. To evaluate this effect, imprints of contact areas were compared for the original-type linear hose and for a stiffer linear hose at two different wheel loads (see Figure A-3). From the imprints, he concluded that the stiffness of the linear hose could significantly influence rut depths. However, no quantitative comparisons were made.

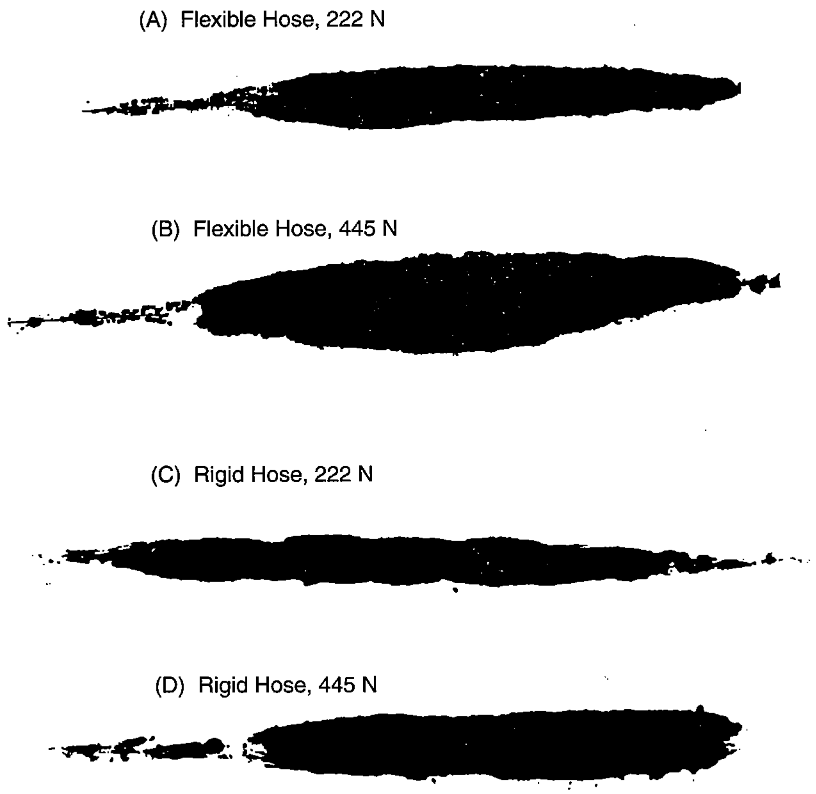


Figure A-3. Imprints of the rubber hose with the beam sample, 689 kPa hose pressure (9).

Also included in this report was a modified testing procedure based upon the improvements to the GLWT and the limited studies. In this procedure, samples were preheated for 6 h in a preheating box to ensure they were heated to the proper temperature. With the new cycle counter, the GLWT would automatically stop running when the desired number of cycles was reached. Finally, the rutting profile of samples was measured using a new template.

Variables identified by this report for possible inclusion within NCHRP Project 9-17 included the following:

- What type of linear hose should be standardized? A stiffer or larger hose than is currently used in the APA might better predict rutting potential.
- How long should samples be preheated? Lai indicated that 6 h in the preheating box was sufficient.

West, Page, and Murphy (*10*) summarized a Florida DOT (FDOT) study performed to (1) verify that the GLWT can be used to predict actual field performance, (2) compare GLWT results with Marshall properties and with gyratory shear properties using the U.S. Army Corps of Engineer's Gyratory Testing Machine, and (3) evaluate the potential of the GLWT to be used as a "proof test" during mix designs.

Laboratory testing for this project consisted of reproducing mixtures to match particular in-place properties from three field projects with known rutting performance. One mix had a very good rutting performance history, one a poor history, and the third a moderate history. Aggregates used for the study were obtained from original sources. Gradations were based on aggregate cores from the three projects. A standard FDOT AC-20 was used in all mixtures as the original asphalt binders were not available.

Compaction of beam samples was accomplished by loading across the top of the beam to 267 kN (60,000 lb), then releasing the load four times. On the fifth application, the 267-kN load was held for 6 min. The compacted beams were allowed to cure at room temperature for 7 days prior to testing. After curing, a conditioning time of 24 h at 40.6°C (105°F) was used. GWLT test conditions were a hose pressure of 690 kPa (100 psi), a wheel load of 543 N (122 lb), and a test temperature of 40.6°C (105°F). Rut depths were obtained at 0; 1,000; 4,000; and 8,000 cycles at seven locations across the top of the beams using the template developed by Lai (*6*).

After testing the beams, West, Page, and Murphy (*10*) noted that rut depth measurements were not consistent across the beam surface. They suggested that the rut depth should be an average of the three center rut depth measurements of the template. More variability existed in rut depth measurements obtained from the outside measuring locations.

The authors concluded that there was a good correlation between the rutting predicted by the GLWT and the actual rutting performance of the mixtures. Compared with Marshall properties, the GLWT provided a better estimate of rutting potential. The authors also suggested that rut depths measured after 1,000 cycles may be used to differentiate the rutting potential of mixtures instead of carrying out the test to 8,000 cycles.

The authors also noted possible drawbacks to the GLWT that may also be applicable to the APA. The first are the scale effects: the small contact area provided by the linear hose and beam samples relative to the size of aggregates within a mixture is not representative of contact areas of tires on highway pavements. The second is uniformity of loadings: the authors surmised that because of the variability in rut depths at the outside measuring locations, the load applied by the linear hose and wheel may not be uniform. Finally, the method of compacting beam samples used during this study may not adequately simulate compaction of pavements in the field. This could affect the aggregate particle alignment and homogeneity of density and will, in turn, affect rutting potential.

Variables identified by this report for possible inclusion within NCHRP Project 9-17 included the following:

- What diameter linear hose should be used? The authors noted that the contact area of the linear hose relative to aggregate sizes is not representative of field conditions.
- Evaluation of rut depth variability across specimens: for acceptance criteria, should rut depths be averaged from the center portion of the specimen or should all measurements be averaged?
- How should specimens be compacted?
- What test temperature should be employed?
- Should specimens be conditioned prior to testing? If so, for how long and at what temperature?
- Should tests be conducted to 8,000 cycles, or are 1,000 cycles sufficient to differentiate between mixtures?

Lai (*11*) presented the results of a round-robin study conducted to evaluate the GDOT standard test procedure for the GLWT. Results of this round-robin study provided between-laboratory and within-laboratory statistics for both the compaction of beams (bulk density) and measured rut depths.

For the round-robin study, beams were compacted using a compression machine. The procedure included heating pre-batched aggregates to 193°C (380°F). Asphalt binder heated to 165.5°C (350°F) was then added to the aggregates, and the combined constituents were mixed. Mixture compaction temperature was 149°C to 154°C. Heated mixture was then placed into 75 mm × 75 mm × 375 mm (3 in. × 3 in. × 15 in.) beam molds, spread, and spaded until the entire mixture was introduced. Compaction consisted of placing a steel plate over the top of the mixture and applying a compressive load. The actual compressive load was not reported; however, it is assumed that the compressive load was applied until a specimen achieved the appropriate height of 75 mm.

GLWT testing was conducted using a test temperature of 40.5°C (105°F) and a hose pressure of 689 kPa (100 psi). Although not reported, it is assumed that a wheel load of 445 N (100 lb) was also used. The loaded wheel tracked on the pressurized hose at a rate of 45 cycles per minute, and rut depths were measured at 0; 500; 1,000; 4,000; and 8,000 cycles.

Standard deviations for the beam densities within and between laboratories were 7.6 and 33.9 kg/m^3, respectively. An average bulk density of 2336 kg/m^3 was reported. The standard deviation on measured rut depths within the laboratory at 8,000 cycles was 0.4 mm with a mean rut depth of 3.4 mm. The between-laboratory standard deviation on measured rut depths was 1.28 mm. The poor reproducibility between laboratories was judged to be caused by the large deviations in the beam bulk densities.

This paper also presented another possible reason for the variation in rut depth measurements. One of the participating laboratories determined the temperature within the GLWT at the beginning of each test and the temperature at the end of the test. Based on a weighted average of temperatures, the rut depths seemed to decrease as the weighted average temperature decreased (see Table A-2).

Variables identified from this paper for possible investigation in NCHRP Project 9-17 included the following:

- What method of compacting test specimens should be employed? The method of compaction for this study was a compressive load. Other studies have used different methods of compaction.
- What is the variability in test parameters during operation? From this study, variations in temperature were identified.

Lai and Shami (*12*) described the development of a compaction device that was incorporated in GDOT's test procedure GDT-115: Method of Test for Determining Rutting Susceptibility Using the Loaded Wheel Tester. The paper reviewed alternative methods of compacting beam specimens for testing in LWTs. Methods such as the kneading compaction, static compaction, and French sample compaction method are discussed. Lai and Shami then discussed the development of the rolling compaction machine. They con-

TABLE A-2 Rut depth versus weighted test temperatures (*11*)

Test No.	Weighted Temp, °C	Rut-Depth (cm) @ 8,000 Cycles
Test 1	43.5	0.343
Test 6	42.4	0.366
Test 3	40.9	0.368
Test 2	40.2	0.323
Test 4	39.1	0.274
Test 5	37.7	0.216

cluded that the rolling compaction machine was easier to use, fabricated beam samples with uniform and controllable densities, and slightly improved the productivity of producing beam samples over the static compaction method.

Collins, Watson, and Campbell (*13*) provided a general overview of the development of the GLWT. It included information from several in-house studies conducted by GDOT that were not previously published. One study evaluated the effects of increased temperature, load, and grade of asphalt binder on GLWT results. Three grades of asphalt binder (AC-10, AC-20, and AC-30), two test temperatures (40.6 and 50°C or 105 and 122°F), and two hose pressures (690 and 827 kPa or 100 and 120 psi) were used. Results showed that all three factors affected measured rut depths, with the test temperature having the most pronounced effect. The authors indicated that increases in tire pressures (or hose pressures) have a direct effect on pavement rutting, but not as much as increases in pavement (or test) temperature.

Another in-house study compared asphalt binders (AC-20s and AC-30s) commonly used in Georgia. Test temperatures of 40.6 and 50°C were used in this study. GDOT had the National Center for Asphalt Technology (NCAT) perform Superpave binder gradings for each binder. All graded as PG 64-22. Interestingly, the GLWT results agreed with those of dynamic shear rheometer testing on rolling thin film oven–aged binders. As $G^*/\sin\delta$ values increased (indicating increased binder stiffness), the measured rut depths generally decreased.

Another in-house study included an experiment to compare SMA with conventional dense-graded mixtures using the GLWT. Results of this study indicated that SMA exhibited approximately half the measured rut depths as the conventional dense-graded mixes.

Potential variables brought out by this paper included the following:

- What test temperature should be specified? The authors used test temperatures that were in excess of the average summer air temperatures, as Lai had done previously (35°C). Using the higher temperature (50°C), the GLWT was able to differentiate between mixtures. Also the authors indicated that temperature may play a more important role in the GLWT than does hose pressure.
- What hose pressure should be specified? Again, the increased hose pressure of 827 kPa was able to differentiate between mixes.

Miller, Ksaibati, and Farrar (*14*) described a research study that evaluated the ability of the GLWT to predict rutting of HMA pavements. Cores having a 150-mm (6-in.) diameter were obtained from several test sites throughout Wyoming, as well as field rut depths and traffic data, and rut tests were preformed with the GLWT on the cores.

The first step in the study was to modify the GLWT to test 150-mm-diameter cores. In order to evaluate whether the modified device would work, the authors performed a repeatability experiment. Twenty-two identical laboratory-prepared cylindrical specimens were fabricated with a combination of kneading and static compaction efforts and then tested at 46.1°C (115°F) to 8,000 cycles. From the experiment, Miller, Ksaibati, and Farrar concluded that the repeatability of the modified GLWT was acceptable.

Next, the authors obtained cores from 13 pavement sections throughout Wyoming. The sections were selected based upon geographic location and field rut performance. For each pavement section, field rut depths, pavement ages, project elevations, EDLAs, highest monthly mean temperature, and type of surface course treatment were all obtained. Three cores from each pavement section were taken from between the wheel paths. Each core was cut to a height of 80 mm.

Cores were tested at both 40.6 and 46.1°C (105 and 115°F). After testing, the authors made an effort to correlate rut test results with actual field rutting using numerous regression models. However, correlations were not strong. Therefore, the authors split the data into two different categories.

The first category was based upon the elevation of the pavement sections. Using the regression analysis containing pavement sections at elevations between 1,158 and 1,676 m (3,800 and 5,500 ft), an R^2 value of 92.6 was achieved. Independent variables in the equation were the average rut depth after 8,000 cycles in the GLWT at 46.1°C and the height of the field core tested. (The use of core height is interesting because the authors indicated in the paper that each core was cut to a height of 80 mm. However, data presented in the paper indicated that core heights ranged from 71 to 84 mm.) A similar regression analysis was also performed on pavement sections at elevations between 1,676 and 2,316 m (5,500 and 7,600 ft). This regression analysis produced an R^2 value of 91.9 with independent variables as the average rut depth after 8,000 cycles at 46.1°C and the center laboratory rut depth after 8,000 cycles at 46.1°C where the center laboratory rut depth was presumably the rut depth at the center of the core specimen.

The second category used for analysis was the existence of a surface treatment on the cores. Surface treatment was defined as either a wearing course (assumed a dense-graded mix) or chip seal. For cores without surface treatments, the regression analysis yielded an R^2 value of 97.3. Independent variables included in this analysis were the average rut depth after 8,000 cycles at 46.1°C and the height of the field core. The regression analysis for the cores with surface treatments produced an R^2 of 93.4 with independent variables of the average rut depth after 8,000 cycles at 46.1°C and the center rut depth after 8,000 cycles at 46.1°C.

From this study, Miller, Ksaibati, and Farrar concluded that the GLWT can be used to test pavement cores instead of beams because the repeatability was acceptable. Rut depths in the GLWT when tested at 40.6°C did not correlate well with field rut depths, but the rut depths for specimens tested at 46.1°C did show correlation with field performance when factors such as elevation and surface treatment were taken into account.

This study raised several questions that could be incorporated into NCHRP Project 9-17 as test variables, including the following:

- What test temperature should be used in the test procedure? The authors concluded that the higher test temperature of 46.1°C showed more promise for predicting field performance.
- One of the primary objectives of NCHRP Project 9-17 is to determine whether the APA can predict field performance. This paper identifies several pieces of information that should be obtained from field pavement sections used in NCHRP Project 9-17 such as elevation, traffic (EDLAs), age of pavement, geographic locations, highest monthly mean temperature, and type of surface treatment.
- Where should rut depth measurements be obtained on a sample in the APA? Rut depth measurements obtained from the center of samples did show significance in some of the regression analysis.
- In order to compare laboratory rutting and field rutting, should NCHRP Project 9-17 use field cut slabs/cores versus beams/cylinders fabricated using original materials from various projects?
- Can different specimen heights be specified?

Collins, Shami, and Lai (*15*) presented the results of a study to evaluate the use of samples compacted with the SGC in the GLWT. Test specimens used in the GLWT prior to this study were predominantly beams. Therefore, the primary modifications to the testing method for SGC cylindrical specimens consisted of developing a holder in which the samples could reside during testing.

The first sample holder evaluated was an aluminum block into which two 150-mm-diameter holes were cut. This holder was not used because of problems extruding samples after testing (especially at higher test temperatures). The second design was a high-density polyethylene holder that was split down the middle. After testing, the two sides of the holder could be separated to remove the sample. This second type was selected.

The test plan for this study was designed to evaluate whether rut depths measured from the SGC-compacted specimens compared with beam samples. Three mixture types were included in this comparison, each with different perceived rutting susceptibility based upon experience. SGC samples were compacted to air void contents of 2, 4.5, and 7 percent. The specimens were compacted by adjusting the number of gyrations to achieve a height of 75 mm and the desired air void content. Beam samples were compacted to 4-, 4.5-, and 5-percent air voids using a rolling wheel compactor. Testing was conducted for the different samples at 40°, 50°, and 60°C. An applied wheel load of 445 N (100 lb) and hose pressure of 689 kPa (100 psi) were used for all testing.

Results of the study indicated that the samples compacted with the SGC correlated well with the standard beam samples (see Figure A-4). Additionally, results indicated that air void contents of the samples and the test temperature also influenced final rut depths. As air voids and temperature increased, so did rut depths.

This paper identified several possible variables to be included within NCHRP Project 9-17:

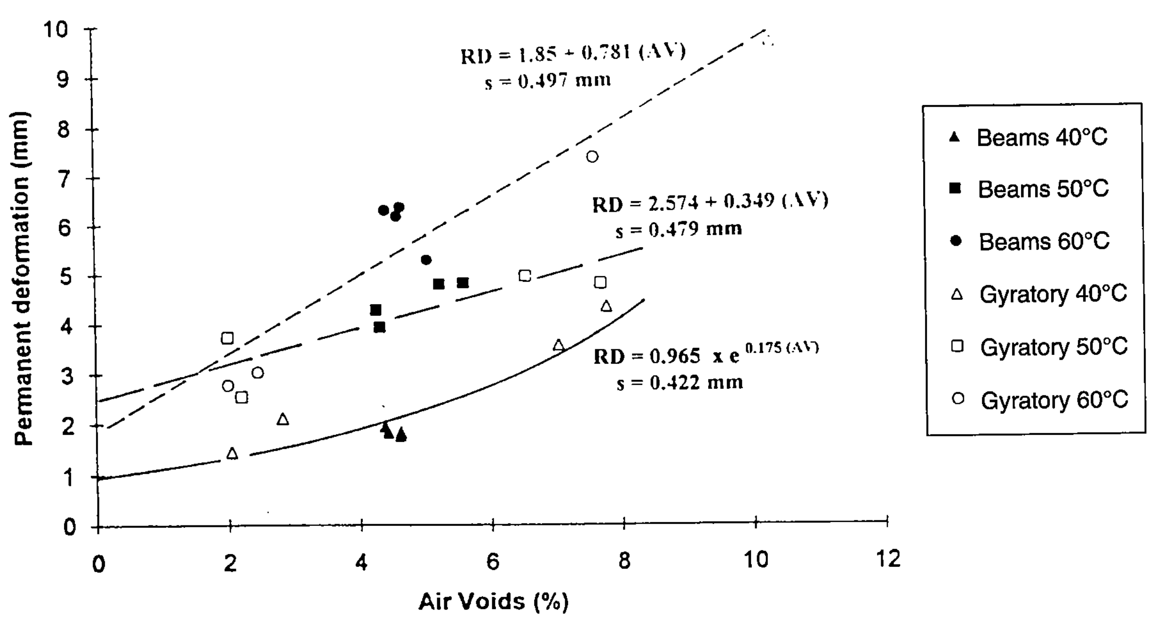


Figure A-4. Rut depth versus air voids of beam samples and gyratory samples (15).

- At what air void content should specimens be tested?
- What test temperature should be used during testing?
- What compaction method should be employed?
- If SGC samples are acceptable, what methodology should be recommended in a standard specification to compact the samples? Compaction of SGC samples in this study was conducted by adjusting the number of gyrations to obtain the required height and density. Some have set the required height and adjusted the mixture mass to obtain the required density while others have compacted normal SGC mix design samples (115 mm) and cut the samples to a height of 75 mm.

An internal GDOT memorandum by Wu (*16*) describes an evaluation of average contact pressures applied by the linear hose in both the GLWT and APA. The author describes several methods of measuring contact pressures, such as triaxial load pin array, pressure-sensing film, and carbon paper. In one experiment, contact pressures measured using the pressure-sensing film and carbon paper were compared. The author concluded that the carbon paper method proved consistent with the pressure-sensing film and therefore selected the carbon paper to conduct an additional experiment.

The second experiment consisted of evaluating the average contact areas for various wheel load/hose pressure combinations. Wheel loads from 359 to 534 N (80 to 120 lb) and hose pressures from 552 to 827 kPa (80 to120 psi) were evaluated. From the results of this experiment, Wu concluded that average contact pressures in both the GLWT and APA are not sensitive to load or hose pressure as long as the wheel load is between 400 and 489 N (90 and 110 lb) and the hose pressure is between 621 and 758 kPa (90 and 110 psi). This was based on similar contact areas (and therefore contact pressures) under these conditions. Wu also concluded that contact pressures in the GLWT were higher than those in the APA under similar hose pressure and wheel loadings (see Tables A-3 and A-4). Based upon the findings of this report, the contact pressure applied to a sample, whether beam or cylindrical, should be examined within NCHRP Project 9-17.

Shami et al. (*17*) described a study to develop a method of predicting rut depth values of an HMA mix with the GLWT for different temperatures and number of loading cycles. For

TABLE A-3 ASTEC APA (contact area and pressure data) (*16*)

	Hose Pressure (kPa)	Contact Area (cm^2)	Contact Pressure (kPa)
Load = 356 N	552	6.13	579
	621	6.13	579
	689	6.00	592
	758	6.00	592
	827	5.81	613
Load = 400 N	552	6.32	633
	621	6.13	655
	689	6.13	655
	758	6.00	668
	827	6.00	668
Load = 445 N	552	6.77	655
	621	6.64	668
	689	6.45	689
	758	6.45	689
	827	6.32	703
Load = 489 N	552	7.29	688
	621	7.10	689
	689	6.97	703
	758	6.97	703
	827	6.97	723
Load = 534 N	552	7.93	675
	621	7.61	703
	689	7.42	717
	758	7.42	717
	827	7.29	730

TABLE A-4 GWLT (contact area and pressure data) (*16*)

	Hose Pressure (kPa)	Contact Area (cm^2)	Contact Pressure (kPa)
Load = 356 N	552	5.35	661
	621	5.16	689
	689	5.03	710
	758	4.84	737
	827	4.84	737
Load = 400 N	552	5.68	703
	621	5.48	730
	689	5.48	730
	758	5.35	744
	827	5.16	779
Load = 445 N	552	6.13	723
	621	6.00	744
	689	6.00	744
	758	5.81	765
	827	5.68	785
Load = 489 N	552	6.64	737
	621	6.45	758
	689	6.45	758
	758	6.13	799
	827	6.00	813
Load = 534 N	552	6.97	765
	621	6.64	806
	689	6.00	889
	758	5.81	916
	827	5.81	916

this study, a total of seven different mixtures were included: five dense-graded and two SMA mixtures. All seven used the same neat AC-30 asphalt binder. Beam samples (125 mm × 300 mm × 75 mm) were compacted to 3- to 5-percent air voids using a rolling wheel compaction machine. Rut tests were conducted at temperatures of 40, 50, and 60°C (104, 122, and 140°F) for three of the mixes and 40 and 60°C for the remaining four mixtures. The applied wheel load was 445 N (100 lb) and the hose pressure was 689 kPa (100 psi). Rut depth measurements were obtained at 0; 1,000; 2,000; 4,000; and 8,000 cycles. The test temperature of 40°C was selected because of the then GDOT test procedure for the GLWT, which represented a typical high air temperature in Georgia. A 60°C test temperature was selected because it represented a high pavement temperature during summer in Georgia.

Results of the rut testing were regressed to obtain best fits for the data. The resulting equation (Equation A.1) allowed for predicting rut depths at various test temperatures, the number of cycles, or both. Shami et al. reported very good correlations (R^2 from 87 to 90 percent) using the following regression equation:

$$\left[\frac{R}{R_o}\right]=\left[\frac{T}{T_o}\right]^{\alpha}\left[\frac{N}{N_o}\right]^{\beta} \qquad \text{(A.1)}$$

where

R = predicted rut depth;
R_o = reference rut depth obtained from the LWT test at the reference conditions T_o and N_o;
T, N = temperature and number of load cycles at which the rut depth is measured;
T_o, N_o = reference temperature and load cycles at the R_o; and
α, β = statistically determined coefficients.

Shami et al. identified two major potential uses for the temperature effect model. First, the testing time for the APA could be reduced for acceptance testing. By testing to 1,000 cycles and predicting rut depths to 8,000 cycles, testing time could be reduced by approximately 2 h. Secondly, the rutting behavior of a particular mix could be predicted for different service conditions and pavement temperatures.

This paper also provides some possible factors for the NCHRP Project 9-17 study:

- What test temperature should be used? The authors discussed both air and pavement temperatures for selection of a test temperature.
- What terminal number of cycles is needed?
- This paper brings out an interesting possibility for the field acceptance of HMA mix. By testing at a higher temperature and a fewer number of cycles, a similar temperature effect model may reduce testing time significantly, thereby increasing the potential for QC/QA applications.

A 1998 Missouri DOT report (*18*) documented a study to evaluate the rutting characteristics of HMA mixes used on high-traffic-volume roadways. The research was conducted to determine whether the GLWT could be used as a laboratory proof test to identify rutting potential. Within the study, the researchers used 40 conventional dense-graded mixtures that were designed using 75 blows of the Marshall hammer, 9 SMA mixes designed by 50 blows of the Marshall hammer, and 14 Superpave-designed mixes.

For testing in the GLWT, the dense-graded and SMA mixes were compacted with the Georgia Rolling Wheel Compactor while the Superpave mixtures were compacted with the SGC. (Target air voids were not reported; however, within an appendix, air void contents for the conventional dense-graded mixtures ranged from 3.8 to 4.8 percent, while the air void contents for the SMA mixes ranged from 4.1 to 4.5 percent. No air void contents were provided for the Superpave gyratory specimens.)

Two test temperatures were used for the dense-graded and SMA mixes: 40 and 60°C (104 and 140°F). The lower temperature is that which is called for in the GDOT test method GDT-115 used at the time of this study; the higher temperature is that which is determined by the Strategic Highway Research Program (SHRP) Superpave pavement temperature model. This model suggested that pavement temperatures during hot summer days in Missouri were in the range of 60 to 70°C. These samples were subjected to 8,000 cycles. The Superpave designed mixes were subjected to a more severe test. The compacted specimens were first vacuum saturated; this was followed by one freeze-thaw cycle and one warm-water soaking cycle. The specimens were then tested at 60°C for 20,000 cycles. This was done to investigate rutting and stripping potential. The data were analyzed similar to the Hamburg LWT in that the post compaction consolidation, creep and stripping slopes, and the stripping inflection point were all determined.

With the test results, the researchers compared rut depths with various mixture properties. For the two dense-graded mix types (75-blow Marshall and Superpave), increases in natural sand content resulted in higher rut depths. Also, as the percentage of aggregate passing the No. 50 sieve increased, the data indicated decreasing rutting potential. Similarly, as the percentage of aggregate passing the No. 100 sieve increased, rutting potential also decreased. Since different asphalt binders were used in each type mixture, the researchers evaluated rut depths versus $G^*/\sin\delta$. The data indicated that as $G^*/\sin\delta$ increased, rutting potential decreased. Testing conducted on binder mixes (larger nominal maximum aggregate size [NMAS] mixtures, 100 percent passing 25.0-mm sieve) was not as clear. Rut depths increased with increasing percentages of natural sand and decreased with increasing $G^*/\sin\delta$ values. For the SMA mixes, only the $G^*/\sin\delta$ parameter was related to rut depths in the expected manner.

From this study, the researchers concluded that the GLWT could be used as a laboratory proof tester. Several questions were raised by this report:

- Should the test temperature for a LWT be related to pavement temperature during hot summer days?
- Can the APA be used to differentiate between binder/base mixes and surface mixes?
- Does the NMAS of the mixture influence test results?

Choubane, Page, and Musselman (*19*) reported the results of an FDOT study to evaluate the APA for assessing the rutting potential of HMA mixes. The test plan for this study included HMA mixtures from three Interstate projects that had been placed during the 1980s. These three mixes are identical to those used in West, Page, and Murphy (*10*). Cores were obtained from the different Interstate projects to determine in-place gradation and asphalt content. Aggregates from the original source were obtained and recombined to meet average gradations determined from the cores. The original asphalt binder was not available, so a standard AC-20 that met original project specifications was used in the mixes.

For each of the three HMA mixtures, the authors compacted nine beams using the Asphalt Vibratory Compactor (AVC). Additionally, 18 cylinders were compacted using the SGC. All testing specimens were targeted for 7-percent air voids. After compaction, all specimens were allowed to cure at room temperature for 7 days. Prior to testing, the specimens were conditioned in the APA at 41°C (105°F) for 24 h. The wheel load and hose pressure used during testing were 540 N (122 lb) and 689 kPa (100 psi), respectively. Rut depth measurements were obtained at 0; 1,000; 4,000; and 8,000 cycles.

Results of the testing showed a good correlation between the average rut depths obtained on the beam and cylindrical specimens. However, the magnitude of the rut depths between the two sample types were statistically different. Based on a linear regression of rut depths between the two sample types, the authors suggested that under similar testing conditions, the cylindrical specimens rutted more when rut depths were below 10 mm, but the beams tended to rut more when rut depths were greater than 10 mm (see Figure A-5).

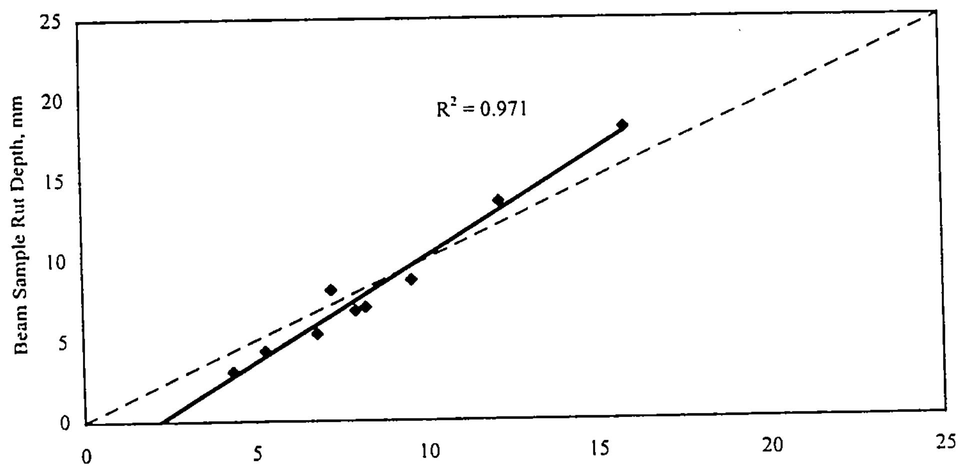


Figure A-5. Comparison of beams and gyratory rut depths (19).

The APA ranked the three mixtures similar to the ranking of the field performance data. The ranking was similar for both the beam and cylindrical specimens. Therefore, the authors concluded that the APA has the potential to rank mixtures for rutting potential. They also expressed concern that the variability both between tests and within each test was relatively high in the APA. For this reason, the authors recommended further evaluation.

From this review, potential variables for inclusion within NCHRP Project 9-17 were identified:

- Comparison between beam and cylindrical specimens.
- Variability both between tests and within each test.

Hanson and Cooley (*20*) reported the results of a research project conducted by NCAT for the South Carolina DOT (SCDOT). It included work with the APA to characterize HMA mixtures of various aggregate gradations and asphalt binder types. Testing conditions in the APA were as follows: test temperature of 64°C (147°F), 445-N (100-lb) wheel load, 689-kPa (100-psi) hose pressure, and 8,000 cycles. Results from the study indicated that the APA was able to differentiate between polymer-modified (PG 76-22) and neat (PG 64-22) asphalt binders. Rut depths were significantly less for mixtures containing polymer-modified binders. Based upon this study and others, the APA has shown the ability to differentiate between asphalt binder types (performance grade).

Prowell (*21*) documented a study conducted at the Virginia Transportation Research Council to evaluate the APA as a tool for estimating the rut susceptibility of HMA. The study consisted of testing numerous field-produced HMA mixes representing a wide range of mix components and types. From the developed database of rut testing, confidence limits were developed and used to evaluate the rut susceptibility of HMA mixes commonly used in Virginia. For the study, the author used a test temperature of 49°C (120°F), a wheel load of 533 N (120 lb), and a hose pressure of 830 kPa (120 psi) for testing beam samples compacted to 7-percent target air voids with an AVC.

Through 1998, a total of 187 HMA mixtures representing 13 different mix types were tested in the APA, and means, standard deviations, and 95-percent confidence limits for the six different mix types were calculated. Using the upper limit of the 95-percent confidence band, criteria for limiting rut depths were proposed. Suggested rut depth criteria for six different mix types ranged from 3.5 mm (a polymer-modified PG 76-22 with a 9.5- or 12.5-mm NMAS 50-blow Marshall designed at 4.5-percent air voids) to 8.5 mm (a fine-graded 9.5-mm NMAS 50-blow Marshall mix designed at 6-percent air voids).

Interestingly, three of the six mixes for which criteria were developed had the same gradation requirements, but different specifications for the asphalt binder grade (SM-2A, SM-2D, and SM-2E). The standard mix (SM-2A) used a PG 64-22 binder while the SM-2D included a one-performance-grade "bump" (PG 70-22) and the SM-2E used a two-grade bump (PG 76-22). Criteria for these three mixes suggested that the addition of the stiffer binders reduced the rut susceptibility of the gradation. The SM-2A proposed rut depth criteria was 7.0 mm; the SM-2D and SM-2E criteria decreased as the performance grade increased (5.5 and 3.5 mm, respectively). This follows the same trend as in previous research.

Based on the study, Prowell indicated that the Virginia DOT (VDOT) would use the APA as a check test to judge the potential performance of new HMA mixes. However, VDOT did not intend to use the APA to predict the rutting of field-placed mixtures. Prowell also suggested that "bumping" the high-temperature performance grade of the asphalt binder resulted in a more rut-resistant and durable pavement. Some concerns expressed in the paper dealt with the beam-to-beam and between-laboratory variability in rut depths for the APA.

Based on this study, several potential factors were identified for the possible inclusion in NCHRP Project 9-17:

- What test temperature, wheel load, and hose pressure should be specified?
- Should HMA mixtures with "bumped" asphalt binder performance grades have the same criteria as mixtures without the bump?

Kandhal and Mallick (*22*) described a research study designed to evaluate the APA as a tool for evaluating the rutting potential of HMA using different aggregate gradations and asphalt binders. Kandhal and Mallick also wanted to develop tentative rut depth criteria based on the work accomplished in the study.

The laboratory test plan included three aggregate types (limestone, granite, and crushed gravel); two asphalt binder types (PG 58-22 and PG 64-22); three gradation shapes (above, below, and through the restricted zone); and two Superpave-defined NMASs (12.5 and 19.0 mm). All factor-level combinations were designed using the Superpave mix design system with a design number of gyrations of 76. Each factor-level combination was then tested in the APA. For testing in the APA, specimens were compacted in the SGC to a height of 75 mm and an air void content of 4 percent. A wheel load of 445 N (100 lb) and a hose pressure of 689 kPa (100 psi) were used in combination with test temperatures of 58 and 64°C (136 and 147°F). The test temperature was based on the asphalt binder used. For those mixes using the PG 58-22 binder, a test temperature of 58°C was used while a test temperature of 64°C was used for mixes with the PG 64-22.

In order to recommend a rut depth criterion, Kandhal and Mallick also obtained mixtures from pavements having major, intermediate, and minor amounts of rutting. These three sections were located on I-85 between Auburn, Alabama, and the Georgia–Alabama state line. Cores were obtained from each section approximately 300 mm (1 ft) from the pavement edge. In the laboratory, the surface mix was cut from the cores and the bulk-specific gravities were determined. Next, the cores were heated and broken apart and recombined for each section. A portion of the reheated mix from each section was then used to determine the theoretical maximum specific gravity while the remaining portion was used to compact 10 specimens for testing in the APA. Specimens were compacted in the SGC to a height of 75 mm and an air void content of 4 percent. The wheel load and hose pressure were identical to that of the laboratory study; a test temperature of 64°C was used.

Results of the laboratory study indicated that the APA was sensitive to aggregate gradation shape (above, below, or through the restricted zone). For the granite and limestone mixes, the gradation passing below the restricted zone (BRZ) showed the highest amount of rutting while the gradation passing through the restricted zone (TRZ) generally showed the least amount of rutting. Interestingly, for the crushed gravel aggregate, the BRZ gradation showed the least amount of rutting; the above the restricted zone (ARZ) gradation showed the highest amount of rutting. The APA was also sensitive to the NMASs of the mixes. Collectively, the 19.0-mm NMAS gradations had lower rut depths. Results also showed significance between the mixes tested at 58 and 64°C. However, Kandhal and Mallick suggested that the significance was probably due to differences in the $G^*/\sin\delta$ value in an unaged condition for the two asphalt binders. The $G^*/\sin\delta$ for the PG 58-22 binder tested at 58°C was 1.27 kPa while the PG 64-22 tested at 64°C had a $G^*/\sin\delta$ of 1.76. This would indicate that at the respective test temperatures, the PG 64-22 would be stiffer.

Based upon the very limited data generated from the three field sections tested, the authors recommended that the APA rut depth after 8,000 cycles should be less than 4.5 mm. The authors recommended that more test sections be evaluated.

This paper brought out several very important variables for consideration in NCHRP Project 9-17:

- What test temperature should be specified?
- Should APA rut depth criteria be different for different NMASs?
- What air void content for APA test specimens should be specified? The authors of this paper used 4 percent as a target air void content; others have used 7 percent.
- What method should be used for testing pavement sections? The project statement for NCHRP Project 9-17 calls for testing of HMA pavement sections with known rutting performance. Within this paper, the authors describe a method of obtaining pavement cores, cutting off the wearing course, heating the cut wearing course, recombining, and then compacting test specimens. Others have obtained original materials and re-fabricated to meet properties of in-place pavements. Both of these methods will have to be considered for NCHRP Project 9-17.

Mallick (*23*) performed a very limited experiment to determine the time needed for a SGC cylinder fabricated for testing in the APA to be heated from room temperature to a test temperature of 64°C (147°F). Mallick used thermocouples embedded into the center of cylinders to measure temperature within the sample. Data from the experiment indicated that after 6 h, the temperature within the sample was 61.7°C (143°F) and did not reach the target temperature until 8 h. This very limited study brought out a very important variable to the successful completion of NCHRP Project 9-17:

- For how long should samples be preheated before testing in the APA?

West (*24*) documented a ruggedness study conducted through the Southeastern Asphalt User Producer Group's APA User Group to identify APA testing factors that have the greatest influence on the outcome of test results. Factors included in the study were air void content of test specimens, specimen preheating time, test temperature, wheel load, hose pressure, and specimen type (cylinder or beam). The specimen type factor actually was a confounded factor in that the

TABLE A-5 Factors and levels used in APA ruggedness study (*24*)

Factors	Low Level Target	High-Level Target
Air Void Content	6 ± 0.5 %	8 ± 0.5 %
Test Temperature	55 ± 0.4°C	60 ± 0.4°C
Specimen Preheat Time	6 h	24 h
Wheel Load	95 lb	105 lb
Hose Pressure	95 psi	105 psi
Specimen Type	AVC beams[1]	SGC cylinders[2]

[1] AVC beams: beams compacted in the Pavement Technology, Inc., Asphalt Vibratory Compactor.
[2] SGC cylinders: cylinders compacted in Superpave Gyratory Compactor.

cylinders were compacted with an SGC and beams were compacted with an Asphalt Vibratory Compactor (AVC). Each factor had two levels, as shown in Table A-5. Another factor that was evaluated was rut potential. This factor was termed a "dummy factor" and was included to represent the range of potential outcomes in the APA; therefore, a low- and high-rut-potential mix were incorporated into the study.

Analysis of the ruggedness study data indicated that three of the six main factors were significant—air void content, test temperature, and specimen type (the dummy factor was also significant)—suggesting that these factors should be more closely controlled than specimen preheat time, wheel load, and hose pressure. Current APA test procedures generally require air void contents to be controlled to within ±1.0 percent. Based on this study, air void contents should be controlled to within ±0.5 percent.

Within the recommendations of this report, West suggests that the APA test chamber temperature, preheating ovens, wheel loads, and hose pressures should be calibrated at least once a year. Also, the sample geometry and compaction method should be specified by individual agencies. A very important recommendation provided in the report is a potential method of identifying outliers in the data. West further suggests that single test results (results for left/right/center or left/right/center/front/back, depending upon testing of beams or cylinders) should not have a standard deviation of greater than or equal to 2.0 mm.

Also included in the report were three areas in which West suggests further work is needed. First, the confounding effect of compactor type and specimen geometry should be studied further. A study of this nature should evaluate whether differences are caused by mold effects, specimen geometries, aggregate particle orientations, or density (air void) gradients. The second recommended study was an evaluation of actual density gradients within samples. This type of study should evaluate the differences in gradients for various sample geometries and compactor types as well as compare density gradients of laboratory-compacted mix with that of field-compacted mix. The final study indicates that more information is needed to provide APA user guidance on how often linear hoses need to be replaced.

This study brought out many important factors that could be considered for NCHRP Project 9-17:

- What tolerance should be used on air void contents of specimens and test temperature? Both of these factors were significant in the ruggedness study.
- What sample geometry and sample compactor should be specified? A possible factor would be to reconfigure cylindrical specimen molds in an effort to yield similar results as beam samples similar to Izzo and Tahmoressi (*4*) and Hall and Williams (*5*). However, the confounding effect mentioned by West may not allow this to happen. Another issue is the repeatability at which testing samples can be produced. Based on experience, it is easier to produce SGC cylindrical specimens at 7.0 ± 0.5 percent air voids than to compact beams with an AVC.
- What type of density gradients occur in specimens fabricated for testing in the APA?
- Do hose pressure and wheel load not affect test results as much as test temperature and air void content do? This question is important because the ultimate goal of the APA is to be able to predict rutting for any application, whether it be for moving traffic or intersections.
- Although not a "factor," a new calibration procedure for the APA may be needed. The report does provide a very strict calibration procedure, which appeared to produce a minimum number of outlier data.

A.4 MISCELLANEOUS

This section describes information found in the literature that does not specifically discuss LWTs, but provides information applicable to NCHRP Project 9-17.

Masad et al. (*25*) describe the distribution of internal air voids within compacted HMA specimens using X-ray tomography. X-ray tomography is a video imaging tool that obtains 3-D images of a solid, bulk medium. Using various algorithms, the images can be used to describe both the components and the distribution of components within the medium. Of interest in this paper is the distribution of air voids throughout the depth of compacted SGC and linear

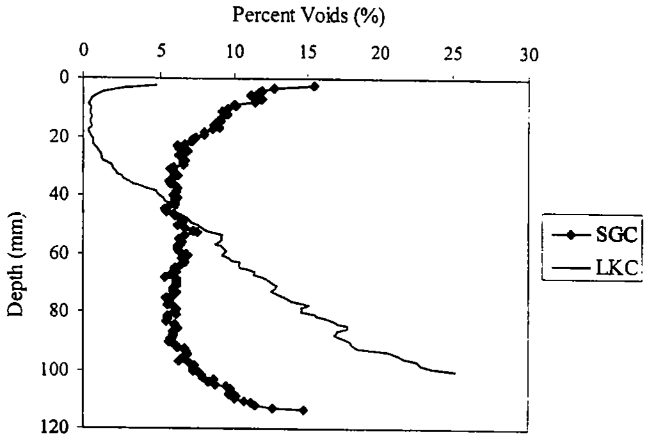


Figure A-6. Void distribution in SGC and LKC specimens (25).

kneading compactor (LKC) specimens. Figure A-6 illustrates these results. For SGC specimens, air voids are typically much higher in the top and bottom 25 mm (1 in.). Within the middle of the specimen, the air void distribution appears constant. However, for the linear kneading compactor specimen, air voids are lower in the top 50 mm (2 in.) and then steadily increase.

A.5 COMPARISONS BETWEEN LWTS

Within this section, information from studies conducted with more than one LWT is provided. The reviews describe the differences in testing conditions used in the studies and in data resulting from the different LWTs.

Stuart and Izzo (*26*) reported results of an FHWA study to validate the Superpave mix design system—namely, the binder tests, binder specifications, and mixture tests. For this study, FHWA used the Accelerated Loading Facility (ALF) located at Turner-Fairbank Highway Research Center. HMA mixes were produced by HMA plants and placed over an aggregate base material on a linear test section. At the time this paper was published, the facility had five surface mixtures and two base mixes in 12 lanes. Each mix type was placed full depth in a lane. The primary objective of this paper was to determine whether the stiffness of asphalt binders as measured by $G^*/\sin\delta$ from the dynamic shear rheometer (DSR) is related to the rut susceptibility of HMA mixtures. A second objective was to determine whether the effects of asphalt binder type on rutting was similar for mixtures with different NMASs.

The research approach taken was to test the seven mixes placed at the ALF for Marshall properties, in different LWTs, and with the U.S. Army Corps of Engineers Gyratory Testing Machine (GTM). Of the seven mixes, five different asphalt binders were used (PG 58-34, PG 58-28, PG 64-22, PG 76-22, and PG 82-22). Each of these binders was combined with aggregates meeting a 19.0-mm NMAS gradation. These mixes were labeled surface mixes. Two of the binders (PG 58-34 and PG 64-22) were combined with aggregates meeting a 37.5-mm NMAS gradation and were designated base mixes.

Response variables included Marshall mix properties; rut depths and rutting slopes using the FRT, GLWT, and HWTD; and the maximum static shear strength, gyratory stability index, gyratory elastoplastic index, and refusal percentage of air voids from the GTM. Target air void contents for all tests were 8 ± 1 percent except for the Marshall properties. The Marshall properties (stability and flow) were determined at 60°C.

The FRT tested two slabs at 60°C using a pneumatic rubber tire. Each slab was 500 mm long × 180 mm wide × 100 mm thick. The rubber tire was 415 mm in diameter with a width of 109 mm. An inflation pressure within the tire was set at 600 kPa, and a load of 5000 N was applied. The wheel-tracking rate for the device was 67 cycles/min (134 passes/min). The two base mixes were not tested with the FRT because the test method is not valid for mixes with NMASs larger than 20 mm. A French plate compactor was used to compact the slabs. This machine uses a reciprocating pneumatic rubber tire to compact the slabs.

The GLWT tested beams with a dimension of 320 mm long × 120 mm wide × 80 mm thick at 40.6°C. The beams were compacted in two lifts using a vibrating hammer with subsequent compaction using a steel wheel until the surface of the beam was flat. Beams were confined by steel plates except for the top 12.7 mm within the GLWT. A stiff rubber hose pressurized to 690 kPa (100 psi) loaded by a steel wheel (700 N or 157 lb) was positioned over the sample and then tracked across the top. Stuart and Izzo noted that the load applied by the wheel was not constant: it changed with the direction of travel. In one direction, the load was approximately 740 N; in the opposite direction, the load was 630 N.

The HWTD tested slabs were submerged in water by rolling a solid steel wheel across the surface of the slab. Slab dimensions were 320 mm long × 260 mm wide × 80 mm thick. Slabs were compacted by the same method as the GLWT beams. The steel wheel had a diameter of 203.5 mm and a width of 47 mm. A load of 710 N was applied by the wheel. The wheel-tracking rate was approximately 53 passes/min.

In order to compare the laboratory properties of the different mixtures to those placed on the ALF, the laboratory mixes were prepared meeting average gradations and asphalt contents of the plant-produced mixes. Marshall properties were then used to verify the lab-produced mixes.

Conclusions from this paper indicated that all three LWTs were able to differentiate among the different asphalt binders in the surface mixes. However, the GLWT and HWTD were not able to differentiate between the surface and base mixes (base mixes were not tested in the FRT). The GLWT provided a very good relationship between rut depths and $G^*/\sin\delta$ while the FRT and HWTD provided reasonably good relationships.

Also of interest in this paper was a side study conducted by the authors. This side study evaluated the conditioning of loose-mix samples prior to compaction of the slabs or beams. Superpave recommends 4 h of oven aging at 135°C (275°F) prior to compaction. The authors aged laboratory mixes for various periods of time (assumed to be at 135°C), extracted and recovered the binder, and performed Superpave binder tests. Mixture from the ALF sections that had been in place for a known amount of time were also cored, the binder extracted and recovered, and the same Superpave binder tests conducted. From their analysis of the data, the authors concluded that a 2-h aging period was needed to produce the average amount of aging that occurred during plant production.

This paper brought out numerous variables that should be considered for inclusion within NCHRP Project 9-17:

- What effect does NMAS have on measured rut depths? Because of the data revealed in this paper, should different rut depth criteria be developed for different NMASs?
- Since all three LWTs were able to differentiate among the different binder types, should test temperatures be based upon the binder performance grade?
- What compaction method should be specified for use with the APA? Two different compaction methods were used in this study.
- How should samples be conditioned prior to testing in the APA? In the side study, the authors suggest that 2 h of oven aging at 135°C prior to compaction may be sufficient.
- What test temperature should be specified? The LWTs used three different temperatures.
- An interesting variable is the rate of wheel tracking. Should the rate be higher or lower than the currently used 60 cycles/min (approximate value) in the APA? Each LWT used different rates of loading.
- What specimen size should be specified? Each LWT used a different sample size. Theoretical boundary conditions and internal stresses should be evaluated.
- Because each LWT used a different contact pressure, what contact pressure should be specified?
- What loading apparatus should be used? A pneumatic tire, pneumatic hose, and steel wheel were used in this study.

Stuart and Mogawer (*27*) continued the work performed by Stuart and Izzo (*26*). Three LWTs were included in this study: FRT, GLWT, and HWTD. This paper presents the results of a study to (1) determine whether the method of specimen compaction affects rutting results in the different LWTs and (2) compare results of rut tests from the LWT testing with data from the ALF. The experimental plan consisted of compacting mixtures with the French Plate Compactor (FPC), an LKC, and pavement slabs cut from sections used in the ALF experiments. Two gradations were used in the study: a surface mixture gradation (19.0-mm NMAS) and a base mix gradation (37.5-mm NMAS). In addition, two asphalt binders were used with both gradations: AC-5 and AC-20 with Superpave performance grades of PG 58-34 and PG 64-22, respectively.

The ALF consisted of a structural frame containing a wheel assembly that models one-half of a single rear truck axle. The wheel traveled at a rate o 1.8 km/h. To simulate highway traffic, the ALF loaded the pavement in only one direction. For the study, the tire pressure was maintained at 690 kPa (100 psi) with a load of 44,500 N (10,000 lb) with no tire wander. During testing, infrared lamps were used to heat the pavement surface to a target temperature of 58°C at a pavement depth of 20 mm. Tests conducted with the FRT, GLWT, and HWTD had identical conditions as those conducted by Stuart and Izzo (*26*).

Conclusions from the study indicated that the ALF provided statistically lower rut depths for the base mixtures. The results varied among devices, compaction method, and mix type (base or surface). Additionally, the results from the LWTs did not always match the ALF results. Stuart and Mogawer hypothesized that the differences could have been caused by differences in contact area for the different tests.

From this paper, several observations can be made. First, all four devices used different types of loading apparatus and conditions. Two used a rubber tire (ALF and FRT), one used a stiff rubber hose (GLWT), and one used a solid steel wheel (HWTD). The ALF and GLWT used inflation pressures of 690 kPa while the FRT used 600 kPa. Loads for the different tests ranged from 44.5 kN for the ALF to 660 N for the HWTD. Additionally, different test temperatures were used for the different devices ranging from 40 to 60°C. Thicknesses also varied for the different devices. Therefore, several potential variables for possible consideration within NCHRP Project 9-17 include the following:

- What type of loading apparatus should be specified?
- What loading conditions should be specified (inflation pressure and wheel load)?
- What sample size should be specified? The authors described potential problems with gradations having large maximum aggregate sizes.
- Are rutting results independent of a gradation's maximum aggregate size (or NMAS)?
- Which contact area and pressure within the APA should be evaluated?
- What test temperature should be specified? Should the specified temperature be relative to high air temperature or high pavement temperature?
- What specimen thickness should be specified?

Huber et al. (*28*) reported on a study that was conducted to evaluate the effect of fine aggregate angularity and flat and elongated particles on the performance of HMA mixes. Three accelerated performance tests were conducted and included the SCRT, APA, and Superpave analysis testing. The SCRT is a Hamburg-type LWT. It uses a single hard rubber tire with an estimated contact pressure of 950 kPa (138 psi) to pass across a beam sample (the beam dimensions were not reported) for 20,000 passes. Specimens tested in the SCRT were compacted to 6.0 ± 0.5 percent air voids using an LKC. Tests were conducted at 55°C (131°F), and rutting was characterized as a rutting rate (mm/h).

Testing with the APA was conducted on 300 mm × 125 mm × 75 mm beams with a hose pressure of 700 kPa (100 psi) and a wheel load of 445 N (100 lb). The beams were compacted to 7.0 ± 1.0 percent air voids using an AVC. Beam specimens were tested at 50°C (122°F), and rutting was characterized as rut depths after 8,000 cycles. The Superpave analysis test used by the authors was the Frequency Sweep at Constant Height Test. This test is a strain-controlled test that applies loads in shear across a sample. Testing was conducted at 40°C.

Two separate experiments were conducted. The first experiment consisted of evaluating the effect of fine aggregates on Superpave-designed mixes. For this study, a 12.5-mm NMAS Superpave mix containing all Georgia granite was used as a reference mix. Three fine aggregates were then used to replace the Georgia granite fine aggregate in order to evaluate the effects of different fine aggregates. Fine aggregate angularity values ranged from 38 to 48 for this experiment. A design number of gyrations (N_{design}) in the SGC of 96 was used. Specimens were compacted to the maximum number of gyrations during mix design, and volumetric properties backcalculated at N_{design}. Results of this experiment indicated that none of the three performance tests were able to differentiate among the four fine aggregates. Huber et al. (*28*) surmised that the gradation combination with the high compactive effort used during mix design had de-emphasized the role of fine aggregates in rut resistance.

The second experiment was conducted to evaluate the effect of flat and elongated particles on volumetric properties of selected asphalt mixtures. No performance testing was conducted in this experiment. Results indicated the coarse aggregate with a maximum to minimum ratio of 3 to 1 did not adversely affect volumetric properties of Superpave-designed mixes. Flat and elongated ratios at 3 to 1 for the two coarse aggregates were 19 and 9 percent.

Based on this study, none of the three performance tests were able to differentiate among fine aggregates with fine aggregate angularity (FAA) values ranging from 38 to 48. This brings out concerns about all three tests. However, testing was conducted for all three performance tests at temperatures that were lower than anticipated pavement temperatures.

Williams and Prowell (*29*) reported on a study conducted jointly by FHWA and the Virginia Transportation Research Council (VTRC) to evaluate the ability of three LWTs to predict the rutting performance of HMA mixtures. The LWTs included were the APA, FRT, and HWTD. For this study, 10 test sections placed in the summer of 1997 at WesTrack were used to correlate LWT results and actual field performance. All 10 of the sections were rehabilitation courses replacing some of the original WesTrack sections. Eight of the ten were part of the FHWA experiment of varying both asphalt content and air voids while the other two sections were Nevada DOT mixes.

Field rut performance on the test sections after 582,000 ESALs was characterized by two permanent deformation components: uplift and downward deformation (see Figure A-7). In the research by Williams and Prowell (*29*), uplift rutting is defined as the difference between total rutting and downward rutting. The authors indicated that none of the LWTs measure

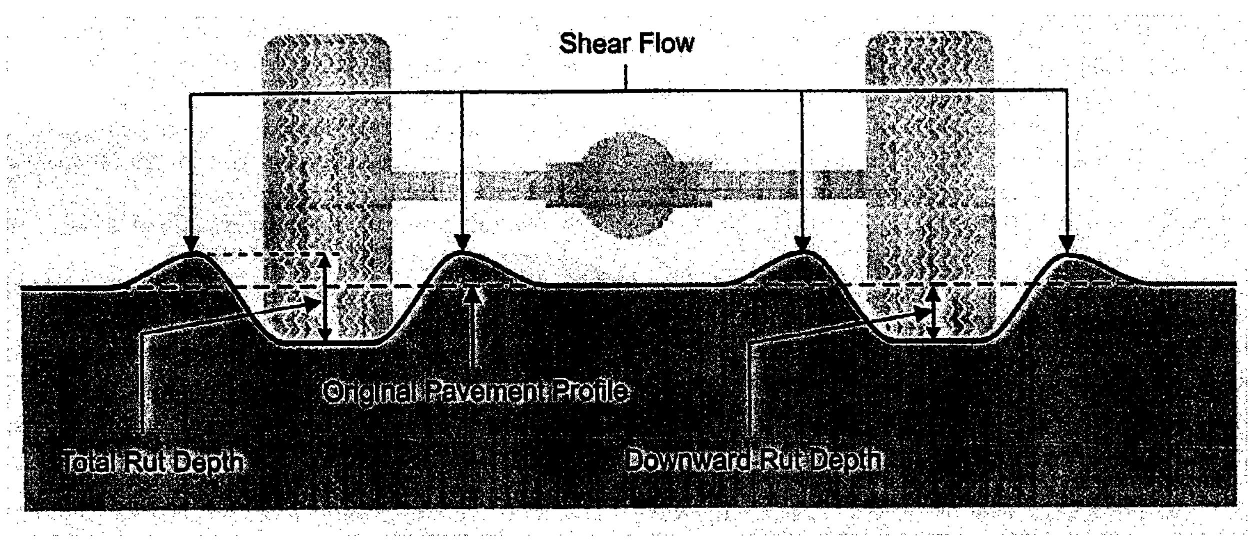


Figure A-7. Characterization of downward and total rutting (29).

TABLE A-6 Test parameters used by Williams and Prowell (*29*)

Parameter	FRT	APA	HWTD
Test Temperature, °C	60	60	50
Environmental Condition	Dry	Dry	Wet
Load Application	Pneumatic tire (600 kPa, 87 psi)	Aluminum wheel on pressurized hose (830 kPa, 120 psi)	Steel wheel
Load, N (lb)	5,000 (1,124)	533 (120)	685 (154)
Wheel Speed	1.6 m/s	0.6 m/s	Sinusoidal with a maximum of 0.33 m/s at the center of sample

the uplift portion of the deformation; they measure only the downward deformation. Field rut performance was obtained from the right wheel path of each test section. The right wheel path was selected because a 2-percent cross slope was used on the test track for drainage; however, depending upon the transverse location in the pavement at which measurements are made, the loading conditions could change. By taking all measurements in the right wheel paths, the LWTs could be compared with the test sections in a consistent manner.

Samples from each of the 10 sections were removed and trimmed to the appropriate thickness for each device. The authors noted that some minor inconsistencies in the samples required the samples to be shimmed in their respective molds using plaster of paris.

Table A-6 presents the testing conditions for each of the three LWTs used by Williams and Prowell. After testing, the results for each device were compared individually with the rut performance data of the 10 pavement sections. The relationships were very strong for all three devices: the APA had an R^2 value of 89.9, and the FRT and HWTD had R^2 values of 83.4 and 90.5, respectively. The paper then presented an approach to establishing specification criteria for LWT devices. The authors suggested that any specification should consider the level of confidence, the variance or standard deviation, sample size, and specification limit.

This paper indicated that a potential exists for correlating LWT results and actual field performance. From this study, several interesting variables were identified for the possible inclusion within NCHRP Project 9-17:

- What is the best method for characterizing field rut depths? The authors suggest that LWTs in general measure only the downward rut depths; therefore, total rut depths such as what LTPP measures may not be correlated to LWT data.
- What test temperature should be specified? Although not discussed in the review, the authors described how the test temperature of 60°C was selected for the FRT and APA. Actual pavement temperature measurements over the time period of the 582,000 ESAL trafficking were used to select the 60°C test temperature.
- What wheel load and hose pressure should be specified for the APA? During this study, the authors used both a wheel load and a hose pressure that were larger than those used in many previous studies and found strong correlations between laboratory and field rut depths.
- Should APA testing be conducted with cores/beams cut from actual pavement sections or should specimens be fabricated in the laboratory to meet section properties?

The paper also describes information that should be considered in developing specification criteria for the APA. This information should be evaluated when developing criteria during NCHRP Project 9-17.

APPENDIX A REFERENCES

1. Nevelt, G., and H. Thanfold. "Evaluation of the Resistance to Deformation of Different Road Structures and Asphalt Mixtures Determined in the Pavement-Rutting Tester." *Proceedings of the Association of Asphalt Paving Technologists*, Vol. 57 (1998); pp. 320–345.
2. Aschenbrener, T. "Comparison of Results Obtained from the French Rutting Tester with Pavements of Known Field Performance." Report No. CDOT-DTD-R-92-11, Colorado DOT (October 1992).
3. Messersmith, P., C. Jones, and C. Wells. "A Contractor's Experience with Polymer Modified Asphalt in Alabama." Paper presented at the 78th Annual Meeting of the Transportation Research Board, Washington, DC (January 10–14, 1999).
4. Izzo, R.P., and M. Tahmoressi. "Use of the Hamburg Wheel Tracking Device for Moisture Susceptibility of Hot-Mix Asphalt." *Transportation Research Record No. 1681*, Transportation Research Board of the National Academies, Washington, DC (1999); pp. 76–85.
5. Hall, K.D., and S.G. Williams. "Effects of Specimen Configuration and Compaction Method on Performance in Asphalt Concrete Wheel-Tracking Tests." Paper prepared for presentation and publication by the Canadian Technical Asphalt Association (June 1, 1999).
6. Lai, J.S. "Development of a Simplified Test Method to Predict Rutting Characteristics of Asphalt Mixes." Final Report, Research Project No. 8503, Georgia DOT (July 1986).

7. Lai, J.S. "Evaluation of Rutting Characteristics of Asphalt Mixes Using Loaded Wheel Tester." Project No. 8609, Georgia DOT (December 1986).
8. Lai, J.S. "Evaluation of the Effect of Gradation of Aggregate on Rutting Characteristics of Asphalt Mixes." Project No. 8706, Georgia DOT (August 1988).
9. Lai, J.S. "Development of a Laboratory Rutting Resistance Testing Method for Asphalt Mixes." Project No. 8717, Georgia DOT (August 1989).
10. West, R.C., G.C. Page, and K.H. Murphy. "Evaluation of the Loaded Wheel Tester." Research Report FL/DOT/SMO/91-391 (December 1991).
11. Lai, J.S. "Results of Round-Robin Test Program to Evaluate Rutting of Asphalt Mixes Using Loaded Wheel Tester." *Transportation Research Record No. 1417,* Transportation Research Board of the National Academies, Washington, DC (1993); pp. 127–134.
12. Lai, J.S., and H. Shami. "Development of Rolling Compaction Machine for Preparation of Asphalt Beam Samples." *Transportation Research Record No. 1492,* Transportation Research Board of the National Academies, Washington, DC (1995); pp. 18–25.
13. Collins, R., D. Watson, and B. Campbell. "Development and Use of the Georgia Loaded Wheel Tester." *Transportation Research Record No. 1492,* Transportation Research Board of the National Academies, Washington, DC (1995); pp. 202–207.
14. Miller, T., K. Ksaibati, and M. Farrar. "Using Georgia Loaded Wheel Tester to Predict Rutting." *Transportation Research Record No. 1473,* Transportation Research Board of the National Academies, Washington, DC (1995); pp. 17–24.
15. Collins, R., H. Shami, and J.S. Lai. "Use of Georgia Loaded Wheel Tester to Evaluate Rutting of Asphalt Samples Prepared by Superpave Gyratory Compactor." *Transportation Research Record No. 1545,* Transportation Research Board of the National Academies, Washington, DC (1996); pp. 161–168.
16. Wu, P. "Contact Pressure Measures of GLWT and ASTEC Asphalt Pavement Analyzer." Interdepartmental Correspondence, Office of Materials and Research, Georgia DOT (July 12, 1996).
17. Shami, H.I., J.S. Lai, J.A. D'Angelo, and T.P. Harmon. "Development of Temperature-Effect Model for Predicting Rutting of Asphalt Mixtures Using Georgia Loaded Wheel Tester." *Transportation Research Record No. 1590,* Transportation Research Board of the National Academies, Washington, DC (1997); pp. 17–22.
18. "Rutting Susceptibility of Bituminous Mixtures by the Georgia Loaded Wheel Tester." Report No. RDT98-001, Missouri DOT (May 1, 1998).
19. Choubane, B., G.C. Page, and J.A. Musselman. "Investigation of the Asphalt Pavement Analyzer for Predicting Pavement Rutting." Research Report FL/DOT/SMO/98-427 (October 1998).
20. Hanson, D.I., and L.A. Cooley, Jr. "Study to Improve Asphalt Mixes in South Carolina, Vol. III: Modified Mixes." Report No. FHWA-SC-98-02 (January 1999).
21. Prowell, B.D. "Development of Rutting Criteria for the Asphalt Pavement Analyzer." Paper submitted for consideration for presentation and publication at the 1999 International Conference on Accelerated Pavement Testing, Paper No. CS6-4 (1999).
22. Kandhal, P.S., and R.B. Mallick. "Evaluation of Asphalt Pavement Analyzer for HMA Mix Design." NCAT Report No. 99-4, National Center for Asphalt Technology (June 1999).
23. Mallick, R.B. Personal Communication (1999).
24. West, R.C. "A Ruggedness Study of the Asphalt Pavement Analyzer Rutting Test." Memorandum to the Asphalt Pavement Analyzer User Group and New APA Owners (May 14, 1999).
25. Masad, E., B. Muhunthan, N. Shashidhar, and T. Harman. "Quantifying Laboratory Compaction Effects on the Internal Structure of Asphalt Concrete." *Transportation Research Record No. 1681,* Transportation Research Board of the National Academies, Washington, DC (1999); pp. 179–185.
26. Stuart, K.D., and R.P. Izzo. "Correlation of Superpave $G^*/\sin\delta$ with Rutting Susceptibility from Laboratory Mixture Tests." *Transportation Research Record No. 1492,* Transportation Research Board of the National Academies, Washington, DC (1995); pp. 176–183.
27. Stuart, K.D., and W.S Mogawer. "Effect of Compaction Method on Rutting Susceptibility Measured by Wheel-Tracking Devices." Paper presented at the 76th Annual Meeting of the Transportation Research Board, Washington, DC (January 12–16, 1997).
28. Huber, G.A., J.C. Jones, P.E. Messersmith, and N.M. Jackson. "Contribution of Fine Aggregate Angularity and Particle Shape to Superpave Mixture Performance." *Transportation Research Record No. 1609,* Transportation Research Board of the National Academies, Washington, DC (1998); pp. 28–35.
29. Williams, C.R., and B.D. Prowell. "Comparison of Laboratory Wheel-Tracking Test Results to WesTrack Performance." Paper presented at the 78th Annual Meeting of the Transportation Research Board, Washington, DC (January 10–14, 1999).

APPENDIX B

PRELIMINARY APA TEST METHOD

Appendix B: Preliminary APA Test Method

(in AASHTO Format)
Standard Test Method for
DETERMINING RUTTING SUSCEPTIBILITY OF ASPHALT PAVING MIXTURES
USING THE ASPHALT PAVEMENT ANALYZER

1. SCOPE

1.1 This method describes a procedure for testing the rutting susceptibility of asphalt-aggregate mixtures using the Asphalt Pavement Analyzer (APA).

1.2 The values stated in SI units are to be regarded as the standard. The values given in parentheses are for information only.

1.3 This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health and determine the applicability of regulations prior to use.

2. REFERENCED DOCUMENTS

2.1 AASHTO Standards

T 166 Standard Test Method for Bulk Specific Gravity and Density of Compacted Bituminous Mixtures Using Saturated Surface–Dry Specimens

T 169 Standard Practice for Sampling Bituminous Paving Mixtures

T 209 Standard Test Method for Theoretical Maximum Specific Gravity and Density of Bituminous Paving Mixtures

T 269 Standard Test Method for Percent Air Voids in Compacted Dense and Open Bituminous Mixtures

MP-2 Standard Specification for Superpave Volumetric Mix Design

PP-35 Practice for Evaluation of Superpave Gyratory Compactors (SGCs)

PP-2 Short and Long Term Aging of Hot Mix Asphalt

3. APPARATUS

3.1 Asphalt Pavement Analyzer (APA)—A thermostatically controlled device designed to test the rutting susceptibility of asphalt-aggregate mixtures by applying repetitive linear loads to compacted test specimens through three pressurized hoses via wheels.

3.1.1 The APA shall be thermostatically controlled to maintain the test temperature and conditioning chamber at any setpoint between 4° and 72°C within 1°C.

3.1.2 The APA shall be capable of independently applying loads up to 534 N (130 lbs) to the three wheels. The loads shall be calibrated to the desired test load by an external force transducer.

3.1.3 The pressure in the test hoses shall be adjustable and capable of maintaining pressure up to 830 kPa (130 psi).

3.1.4 The APA shall be capable of testing six cylindrical or three beam specimens simultaneously.

3.1.5 The APA shall have a programmable master cycle counter which can be preset to the desired number of cycles for a test. The APA shall be capable of automatically stopping the test at the completion of the programmed number of cycles.

3.1.6 The hoses shall be Gates 77B Paint Spray and Chemical 3/4 inch (19.0 mm), 750 psi (5.17 MPa) W.P. GL 07148. The hoses should be replaced when any of the outer rubber casing has worn through and threads are exposed. Follow the APA manufacturer's instructions for the technique on replacing hoses.

3.2 Balance, 12,000 gram capacity, accurate to 0.1 gram.

3.3 Mixing utensils (bowls, spoon, spatula).

3.4 Ovens for heating aggregate and asphalt binder.

3.5 Compaction device and molds.

4. CALIBRATION

4.1 Temperature calibration of the APA and specimen preheating.

4.2 Load calibration of the air cylinders at the three test positions.

4.3 Pressure calibration of the air pressure within APA hoses.

Note: Calibration procedures are provided in the Annex.

5. PREPARATION OF TEST SPECIMENS

5.1 Number of Test Specimens - Six cylindrical (150 mm diameter x 115 or 75 mm) or 3 beam (75 mm x 125 mm x 300 mm) specimens.

5.2 Roadway Core Specimens

5.2.1 Roadway core specimens shall be 150 mm diameter with all surfaces of the perimeter perpendicular to the surface of the core within 5 mm. Cores shall be trimmed with a wet masonry saw to a height of 75 ± 3 mm. If the core has a height of less than 75 ± 3 mm, plaster-of-paris may be used to achieve the proper height. Testing shall be conducted on the uncut face of the core.

5.3 Plant Produced Mixtures

5.3.1 Samples of plant produced mixtures shall be obtained in accordance with AASHTO T 169. Mixture samples shall be reduced to the appropriate test size and compacted while the mixture is still hot. Reheating of loose plant mixture should be avoided.

5.3.2 Mixture samples to be compacted with a Superpave gyratory compactor shall be compacted 4.0 ± 0.5 percent air voids. Beam samples shall be compacted to 5.0 ± 0.5 percent air voids. Both sample types shall be 75 ± 3 mm in height.

5.4 Laboratory Prepared Mixtures

5.4.1 Mixture proportions are batched in accordance to the desired Job Mix Formula.

5.4.2 The temperature to which the asphalt binder must be heated to achieve a viscosity of 170 ± 20 cSt shall be the mixing temperature. For modified asphalt binder use the mixing temperature recommended by the binder manufacturer.

5.4.3 Dry mix aggregates and hydrated lime (when lime is used) first, then add optimum percentage of asphalt cement. Mix the materials until all aggregates are thoroughly coated.

5.4.4 Test samples shall be aged in accordance with the short-term aging procedure in AASHTO PP2.

5.4.5 The temperature to which the asphalt binder must be heated to achieve a viscosity of 290 ± 30 cSt shall be the compaction temperature. For modified asphalt binders, use the compaction temperature recommended by the binder manufacturer. The mixture shall not be heated at the compaction temperature for more than one hour.

5.5 Laboratory Compaction of Specimens

5.5.1 A Superpave gyratory compactor approved in accordance with AASHTO PP-35 should be used to compact cylindrical samples. Beam samples shall be compacted to the proper dimensions and air void content.

5.5.2 Laboratory prepared cylindrical specimens shall be compacted to contain 4.0 ± 0.5 percent air voids at 75 ± 3 mm height.

5.5.3 Laboratory prepared beam specimens shall be compacted to contain 5.0 ± 0.5 percent air voids at 75 ± 3 mm.

5.5.4 Compacted specimens should be left at room temperature (approximately 25°C), to allow the entire specimen to cool, for a minimum of 3 hours.

6. DETERMINING THE AIR VOID CONTENTS

6.1 Determine the bulk specific gravity of the test specimens in accordance with AASHTO T 166.

6.2 Determine the maximum specific gravity of the test mixture in accordance with AASHTO T 209.

6.3 Determine the air void contents of the test specimens in accordance with AASHTO T 269.

7. SELECTING THE TEST TEMPERATURE

7.1 The test temperature shall be set to the high temperature of the standard Superpave binder Performance Grade (PG) identified by the specifying agency for the project that the mixture is intended. For circumstances where the binder grade has been bumped, the APA test temperature will remain at the high standard PG binder.

8. SPECIMEN PREHEATING

8.1 Place the specimens in the molds.

8.2 Specimens shall be preheated in the temperature calibrated APA test chamber or a separate calibrated oven for a minimum of 6 hours. Specimens should not be held at elevated temperatures for more than 24 hours prior to testing.

9. PROCEDURE

9.1 Set the hose pressure gap reading to 827 ± 35 kPa (120 ± 5 psi). Set the load cylinder pressure reading for each wheel to achieve a load of 534 ± 22 N (120 ± 5 lb.).
9.2 Stabilize the testing chamber temperature at the temperature selected in Paragraph 7.
9.3 Secure the preheated, molded specimens in the APA. The preheated APA chamber should not be open more than 6 minutes when securing the test specimens into the machine. Close the chamber doors and allow 10 minutes for the temperature to restabilize prior to starting the test.
9.4 Apply 25 cycles to seat the specimens before the initial measurements. Make adjustments to the hose pressure as needed during the 25 cycles.
9.5 Open the chamber doors, unlock and pull out the sample holding tray.
9.6 Place the rut depth measurement template over the specimen. Make sure that the rut depth measurement template is properly seated and firmly rests on top of the testing mold.
9.7 Zero the digital measuring gauge so that the display shows 0.00 mm with the gauge completely extended. The display should also have a bar below the "inc." position. Take initial readings at each of the four outside locations on the template. The center measurement is not used for cylindrical specimens. Measurements shall be determined by placing the digital measuring gauge in the template slots and sliding the gauge slowly across the each slot. Record the smallest measurement for each location to the nearest 0.01 mm.
9.8 Repeat steps 9.6 and 9.7 for each set of cylinders or beams in the testing position. All measurements shall be completed within six minutes.
9.9 Push the sample holding tray in and secure. Close the chamber doors and allow 10 minutes for the temperature to equalize.
9.10 Set the PRESET COUNTER to 8,000 cycles.
9.11 Start the test. When the test reaches 8,000 cycles, the APA will stop and the load wheels will automatically retract.
9.12 Repeat steps 9.5 through 9.7 after 8,000 cycles to obtain final measurements.

10. CALCULATIONS

10.1 The rut depth at each location is determined by subtracting the final measurement from the initial measurement.
10.2 Determine the overall average rut depth for each of the three test positions. Use the average of measurements to calculate the average rut depth.
10.3 Calculate the average rut depth from the three test positions. Also, calculate the standard deviation for the three test positions.
10.4 Outlier evaluation—If the standard deviation of the three test positions is greater than or equal to 2.0 mm, then the position with the rut depth farthest from the average <u>may</u> be discarded. The testing procedure, device calibration, and test specimens should be investigated to determine the possible causes for the excessive variation.
10.5 The APA rut depth for the mixture is the average of the six cylindrical or 3 beam specimens at 8,000 cycles.

11. REPORT

11.1 The test report shall include the following information:

11.1.1 The laboratory name, technician name, and date of test.

11.1.2 The mixture type and description.

11.1.3 Average air void content of the test specimens.

11.1.4 The test temperature.

11.1.5 The average manually determined rut depth, to the nearest 0.1 mm, after 8,000 cycles.

12. PRECISION AND BIAS

12.1 No work has been conducted to develop a precision statement for this standard.

ANNEX
(Mandatory Information)

A. CALIBRATION

The following items should be checked for calibration no less than once per year: (1) preheating oven, (2) APA temperature, (3) APA wheel load, and (4) APA hose pressure. Instructions for each of these calibration checks is included in this section.

A.1. Temperature calibration of the preheating oven.

A.1.1 The preheating oven must be calibrated with a NIST traceable thermometer (an ASTM 65°C calibrated thermometer is recommended) and a metal thermometer well to avoid rapid heat loss when checking the temperature.

A.1.2 Temperature Stability

A.1.2.1 Set the oven to the chosen temperature (e.g., 64°C). Place the thermometer in the well and place them on the center of the shelf where the samples and molds will be preheated. It usually takes an hour or so for the oven chamber, well and thermometer to stabilize. After one hour, open the oven door and read the thermometer without removing it from the well. Record this temperature. Close the oven door.

A.1.2.2 Thirty minutes after obtaining the first reading obtain another reading of the thermometer. Record this temperature. If the readings from step A.1.2.1 and A.1.2.2 are within 0.4°C, then average the readings. If the readings differ by more than 0.4°C then continue to take readings every thirty minutes until the temperature stabilizes within 0.4°C on two consecutive readings.

A.1.3 Temperature Uniformity

A.1.3.1 To check the uniformity of the temperature in the oven chamber, move the thermometer and well to another location in the oven so that they are on a shelf where samples and molds will be preheated, but as far as possible from the first location. Take and record readings of the thermometer at the second location every thirty minutes until two consecutive readings at the second location are within 0.4°C.

A.1.3.2 Compare the average of the two readings at the first location with the average of the stabilized temperature at the second location. If the average temperatures from the two locations are within 0.4°C, then the oven temperature is relatively uniform and it is suitable for use in preheating APA samples. If the average of the readings at the two locations differ by more than 0.4°C then you must find another oven that will hold this level of uniformity and meets calibration.

A.1.4 Temperature Accuracy

A.1.4.1 Average the temperatures from the two locations. If that average temperature is within 0.4°C of the set point temperature on the oven, then the oven is reasonably accurate and calibration is complete.

A.1.4.2 If the set point differs from the average temperature by more than 0.4°C, then adjust the oven set point appropriately to raise or lower the temperature inside the chamber so that the thermometer and well will be at the desired temperature (e.g., 64°C).

A.1.4.3 Place the thermometer and well in the center of the shelf. At thirty-minute intervals, take readings of the thermometer. When two consecutive readings are within 0.4°C, and the average of the two consecutive readings are within 0.4°C of the desired test temperature (e.g., 64°C), then the oven has been properly adjusted and calibration is complete. If these two conditions are not met, then repeat steps A.1.4.2 and A.1.4.3.

A.2 APA Temperature Calibration

A.2.1 The APA must be calibrated with a NIST traceable thermometer (an ASTM 65°C calibrated thermometer is recommended) and a metal thermometer well to avoid rapid heat loss when checking the temperature.

A.2.2 Temperature Stability

A.2.2.1 Turn on the APA main power and set the chamber temperature controller so that the inside the testing chamber is at anticipated testing temperature (e.g., 64°C). Also, set the water temperature controller to achieve the anticipated testing temperature. (Note: Experience has shown that the temperature controller on the APA is not always accurate. The thermometer should always be considered chamber temperature.) Place the thermometer in the well and place them on the left side of the APA where the samples and molds will be tested (Note: It may be helpful to remove the hose rack from the APA during temperature calibration to avoid breaking the thermometer.)

A.2.2.2 It usually takes about five hours for the APA to stabilize. After the temperature display on the controller has stabilized, open the chamber doors and read the thermometer without removing it from the well. Record this temperature. Close the chamber doors.

A.2.2.3 Thirty minutes after obtaining the first reading obtain another reading of the thermometer. Record this temperature. If the readings from step A.2.2.2 and A.2.2.3 are within 0.4°C, then average the readings. If the readings differ by more than 0.4°C then continue to take readings every thirty minutes until the temperature stabilizes within 0.4°C on two consecutive readings.

A.2.3 Temperature Uniformity

A.2.3.1 To check the uniformity of the temperature in the APA chamber, move the thermometer and well to the right side of the APA, where the samples are tested. Take and record readings of the thermometer at the second location every thirty minutes until two consecutive readings at the second location are within 0.4°C.

A.2.3.2 Compare the average of the two readings obtained in A.2.2.3 and A.2.3.1. If the average temperatures from the two locations are within 0.4°C, then the APA temperature is relatively uniform and it is suitable for use. If the average of the readings at the two locations differ by more than 0.4°C then consult with the manufacturer on improving temperature uniformity.

A.2.4 Temperature Accuracy

A.2.4.1 Average the temperatures from the two locations. If that average temperature is within 0.4°C of the desired test temperature (e.g., 64°C), then the APA temperature is reasonably accurate and calibration is complete,

A.2.4.2 If the average temperature differs from the desired test temperature (e.g., 64°C) by more than 0.4°C, then adjust the APA temperature controller so that the thermometer and well will be at the desired test temperature. (Note: It is advisable to keep the water bath set at the same temperature as the test chamber.)

A.2.4.3 Place the thermometer and well in the center of the shelf. At thirty minute intervals, take readings of the thermometer. When two consecutive readings are within 0.4°C, and the average of the two consecutive readings are within 0.4°C of the desired test temperature, then the APA temperature has been properly adjusted and calibration at that temperature is complete. Record the current set points on the temperature controllers for later reference. If these two conditions are not met, then repeat steps A.2.4.2 and A.2.4.3.

A.3 APA Wheel Load calibration of the air cylinders at the three test positions.

A.3.1 The APA wheel loads will be checked with the calibrated load cell provided with the APA. The loads will be checked and adjusted one at a time while the other wheels are in the down position and bearing on a dummy sample or wooden block of approximately the same height as a test sample. Calibration of the wheel loads should be accomplished with the APA at room temperature. A sheet is provided to record the calibration loads.

A.3.1.1 Remove the hose rack from the APA.

A.3.1.2 Jog the wheel carriage until the wheels are over the center of the sample tray when the wheels are in the down position.

A.3.1.3 Raise and lower the wheels 20 times to heat up the cylinders.

A.3.1.4 Adjust the bar on top of the load cell by screwing it in or out until the total height of the load cell-load bar assembly is 105 mm.

A.3.1.5 Position the load cell under one of the wheels. Place wooden blocks or dummy samples under the other two wheels.

A.3.1.6 Zero the load cell.

A.3.1.7 Lower all wheels by turning the cylinder switch to CAL.

A.3.1.8 If the load cell is not centered left to right beneath the wheel, then raise the wheel and adjust the position of the load cell. To determine if the load cell is centered front to back beneath the wheel, unlock the sample tray and move it SLOWLY until the wheel rests in the indentation on the load cell bar (where the screw is located).

A.3.1.9 After the load cell has been properly centered, adjust the pressure in the cylinder to obtain 445 ± 5 N (100 ± 1lbs.). Allow three minutes for the load cell reading to stabilize between adjustments. Record the pressure and the load.

A.3.1.10 With the wheel on the load cell remaining in the down position, raise and lower the other wheels one time. Allow three minutes for the load cell reading to stabilize. Record the pressure and the load.

A.3.1.11 With the other wheels remaining in the down position, raise and lower the wheel over the load cell. Allow three minutes for the load cell reading to stabilize. Record the pressure and the load.

A.3.1.12 Repeat steps A.3.1.5 through A.3.1.11 for each wheel/cylinder.

A.3.1.13 Return the load cell to the first wheel and repeat steps A.3.1.5 through A.3.1.11.

A.3.1.14 Place the load cell under the second wheel and repeat steps A.3.1.5 through A.3.1.11.

A.3.1.15 Place the load cell under the third wheel and repeat steps A.3.1.5 through A.3.1.11. The current cylinder pressures will be used to set wheel loads to 445 N (100 lbs.).

A.4 Replacement of the APA hoses

A.4.1 New hoses shall be placed in service in accordance with 2.1.6.

A.4.1.1 Remove the hose rack from the APA.

A.4.1.2 Remove the used hoses from the hose rack. Place the new hose on the barbed nipples and secure with the hose clamps.

A.4.1.3 Position the hoses in the rack such that the hose curvature is vertical. Tighten the nuts at the ends of the hoses only until the hoses are secure. Over-tightening will effect the contact pressure and hose life.

A.4.1.4 Place the hose rack back into the APA and make sure that the hoses are aligned beneath the wheels.

A.4.1.5 Prior to testing, break in the new hoses by running 8000 cycles on a set of previously tested samples at a temperature of 55°C (131°F) or higher.

A.5 APA Hose Pressure Check

A.5.1 The air pressure in the APA test hoses shall be checked with a NIST traceable test gauge or transducer with a suitable range. The check shall be made while the APA is operating Since the hoses are connected in series, it is satisfactory to connect the test gauge to the end of the right-most hose. The pressure should not fluctuate outside of the range of 690 ± 35 kPa (100 ± 3 psi) during normal operation. Adjust the pressure as necessary with the hose pressure regulator.

Note: The Ashcroft test gauge model 450182As02L200# has been found to be satisfactory for this purpose. This gauge may be available through Grainger (Stock No. 2F008).

Abbreviations used without definitions in TRB publications:

AASHO	American Association of State Highway Officials
AASHTO	American Association of State Highway and Transportation Officials
APTA	American Public Transportation Association
ASCE	American Society of Civil Engineers
ASME	American Society of Mechanical Engineers
ASTM	American Society for Testing and Materials
ATA	American Trucking Associations
CTAA	Community Transportation Association of America
CTBSSP	Commercial Truck and Bus Safety Synthesis Program
FAA	Federal Aviation Administration
FHWA	Federal Highway Administration
FMCSA	Federal Motor Carrier Safety Administration
FRA	Federal Railroad Administration
FTA	Federal Transit Administration
IEEE	Institute of Electrical and Electronics Engineers
ITE	Institute of Transportation Engineers
NCHRP	National Cooperative Highway Research Program
NCTRP	National Cooperative Transit Research and Development Program
NHTSA	National Highway Traffic Safety Administration
NTSB	National Transportation Safety Board
SAE	Society of Automotive Engineers
TCRP	Transit Cooperative Research Program
TRB	Transportation Research Board
U.S.DOT	United States Department of Transportation

TRANSPORTATION RESEARCH BOARD
500 Fifth Street, NW
Washington, DC 20001

ADDRESS SERVICE REQUESTED