TriDurLE

National Center for Transportation
Infrastructure Durability & Life-Extension

Project ID: 2020-MST_TAM-02

DEVELOPMENT OF HOLISTIC METHODOLOGIES FOR IMPROVING ASPHALT MIX DURABILITY (Yr 1)

Final Report

By

Jenny Liu
Professor
Missouri University of Science and Technology

Fujie Zhou
Senior Research Engineer
Texas A&M University/Texas A&M Transportation Institute

Pedro Romero
Associate Professor
The University of Utah

Yizhuang David Wang
Research Associate
Missouri University of Science and Technology

Bo Lin
Ph.D. student
Missouri University of Science and Technology

for

National University Transportation Center TriDurLE
Department of Civil & Environmental Engineering
405 Spokane Street, PO Box 642910
Washington State University, Pullman, WA 99164-2910

May 2022

Acknowledgements

The authors extend their gratitude to the National Center for Transportation Infrastructure Durability & Life-Extension for funding this research.

Disclaimer

Table of Contents

List of Figures

List of Tables

Executive Summary

Asphalt mix durability has always been one of the major concerns of all highway agencies. To have a durable mix, one needs to address three aspects: durable mix design, production, and placement. The objective of this project is to develop holistic methodologies for addressing all three aspects with an ultimate goal to improve asphalt mix durability. As a minimum, this project will develop a systematic balanced mix design (BMD) methodology for designing durable mixes in the laboratory, recommend a performance-related methodology for production quality control and quality assurance (QC/QA) at asphalt plants, and identify innovative practices for placement acceptance in the field. This report summarizes the findings from the research efforts in Year 1.

During this study, a comprehensive literature review was performed regarding the efforts in ensuring the durability and performance of asphalt mixture in mix design and production. A coherent BMD/QC/QA framework was established that had four interconnected components: (a) volumetric mix design and selection of multiple asphalt contents for mixture performance evaluation, (b) mixture performance evaluation at multiple asphalt contents and selection of the balanced asphalt content, (c) mixture performance verification at the balanced asphalt content, and (d) production QC/QA testing. One set of laboratory tests, namely the ideal cracking test (IDEAL-CT) and ideal rutting test (IDEAL-RT), and associated acceptance criteria were recommended for the mixture performance evaluation and the production QC/QA testing and another set was recommended for the mixture performance verification wherein DOTs could choose their preferred performance verification tests. Furthermore, two practical loose-mixture aging protocols were developed, one for short-term aging was used in the process of the volumetric mix design, the mixture performance evaluation, and the production QC/QA testing, and the other was for mid-term aging employed in the mixture performance verification. The recommended short-term aging protocol was to age the loose mixture in a force draft oven for 2 hours at the mixture compaction temperature while the mid-term aging protocol consists of three steps: (1) short-term aging, (2) 20-hr loose-mixture aging at 100°C, and (3) reheating for compaction. A case study was presented in this report to demonstrate the whole process of the framework, including the actual plant production QC testing.

In addition, the variability of main performance tests in mix design were investigated including semi-circular bend (SCB), Illinois flexibility index (I-FIT), and IDEAL-CT. The inter-laboratory testing variability was also evaluated. The experimental testing matrix included mixtures with two different aggregate gradations, two binders, and three specimen geometries. It was found that both the I-FIT and the IDEAL-CT tests result in approximately the same variability, and no lab was always more (or less) variable than the others. The statistical information should be used to develop a process by which the validity of the data could be assessed based on 3 or 4 samples.

The study also investigated the effects of aging on mixture performance. The samples with different aging conditions, i.e., at the plant, at laydown, and from cores after three years of service, were collected and tested. The testing results were compared with the field performance. All the tests predicted which section would have inadequate performance, but no agreement amongst the different tests for the ‘intermediate’ performing sections. It was not clear if index-type tests, such as the ones performed as part of this research, were meant for such fine-tuned predictions instead of providing pass-fail information.

Chapter 1 Introduction

1.1 Problem Statement

Every year, around 360 million tons of asphalt mixes with a cost of more than $20 billion are placed on the roads in the United States. Asphalt mixes designed with too much asphalt binder may be susceptible to rutting (Figure 1.1 a) while those with too little asphalt binder are prone to cracking (Figure 1.1 b), raveling and other durability related distresses. In the last two decades, state Departments of Transportation (DOTs) have been dealing with durability problem (e.g., cracking) of asphalt mixes.

(a) Rutting (b) Cracking

Figure 1.1 Pavement Major Distresses: (a) Rutting and (b) Cracking

There are many factors causing the durability problem with asphalt mixtures. One of them is the low asphalt content that results from the Superpave volumetric mix design method used by most DOTs. Volumetric design methods in general have three major limitations:

- <u>Subjective calculation of voids in mineral aggregate (VMA):</u> VMA is highly dependent on an accurate measurement of aggregate bulk specific gravity which is both time consuming and subjective. Therefore, VMA verification between two labs is difficult. When the design air voids is fixed, the most critical concern is that the effective asphalt binder content is controlled by the VMA. A questionable VMA calculation often leads to a debatable optimum asphalt binder content. This issue becomes more serious when using reclaimed asphalt pavements (RAP) and recycled asphalt shingles (RAS) in asphalt mixes.
- <u>Poor at evaluating the effect of binder source and modification/dosage selection</u>: The volumetric mix design alone is not able to assess the effects of binder sources, and modifications on mix properties.

Mixes having the same volumetric properties and the same performance-grade binders, but from different sources or modified with various dosages or technologies, may have completely different performance in the field. The influence of those modifications on mix performance is not evident with the volumetric mix design method.

- <u>Inability to predict the effect of recycled materials, rejuvenators, and other additives</u>: Most DOTs allow to the use of recycled materials (e.g.,. 25% binder replacement from RAP/RAS/FRAP for upper layer). But there is no way to know the degree to which the recycled binders are activated and behave as a composite material with the other binder and/or rejuvenators. As the use of recycled materials increases, the amount of material that is not properly characterized in component tests also increases. A test on the end product is the most direct way to go.

To address the durability problem of asphalt mixtures, DOTs have explored different ways to modify the volumetric design method, such as lowering N_{design} or design air voids, increasing VMA requirement, regression N_{design}, regression air voids, etc. All these alternatives may get more asphalt binder into the mixtures and improve the durability of asphalt mixtures. However, as discussed above, the volumetric method alone cannot ensure a good cracking or rut resistant mixture. Thus, there is an urgent need to incorporate performance tests (cracking/rutting/moisture damage) into asphalt mix design and quality assurance process.

There has been a sweeping trend of including simple performance tests as a supplement of the volumetric design in recent years. Those ideas spawn the concepts of balance mix designs (BMDs) / performance-engineered mix designs (PEMDs). Figure 1.2 shows three approaches of implementing the BMD method: (1) Volumetric design with performance testing validation (the left flowchart in Figure 1.2), (2) Performance-modified volumetric mix (the middle flowchart in Figure 1.2), and (3) Performance design (the right flowchart in Figure 1.2). It is worth noting that each approach has its advantages and limitations. Compared to Approaches 2 and 3, Approach 1 is the easiest one to understand and implement, but its limitation is to evaluate one asphalt content. Necessary enhancement to Approach 1 is to test three asphalt contents corresponding to air voids varying from, say 2.5-4%. Approach 2 chooses final asphalt content based solely on performance tests and performance model predictions. The accuracy of the performance tests and performance models becomes extremely important since volumetric properties are not considered. Approach 3 is a kind of “reversed” version of Approach 1. Performance requirements are met first and then the volumetric properties are determined. Implementing Approach 3 requires mindset changes. Thus, it may be better to adopt Approach 3 at a later time after agencies have more experiences on implementing other BMD design approaches. These three approaches have been tried to different extents in the past. In mid 2000s, Zhou et al. (2007) proposed a BMD approach which was established on the fact that a durable mix must have balanced rutting and cracking resistances and later they expanded the BMD to consider specific conditions of the project in terms of traffic volume, climate, mix applications (new construction vs. overlays), and existing pavement conditions for asphalt overlays (Zhou et al. 2014). This research team has also used the volumetric design with performance testing validation approach to construct more than 100 projects in Texas, including stone matrix asphalt (SMA), thin overlay mixes, and RAP/RAS mixes (Zhou et al. 2012, 2014, Im et al. 2016). Other states including Utah, New Jersey, Louisiana, Illinois, and Wisconsin, also constructed different field test sections with Approach 1. California piloted Approach 2 for several high-profile projects; meanwhile Rutgers University of New Jersey has explored Approach 3 in the field. However, none of the three BMD approaches have been fully implemented by any DOT. In summary, the BMD concept has been widely accepted by DOTs. However, there is lack of a systematic and innovative methodology of incorporating all components to achieve a durable mix.

Researchers have spent considerable efforts on developing quality assurance specifications with performance tests. The WesTrack Performance-Related Specification (PRS) project and National Cooperative Highway Research Program (NCHRP) 9-22 project developed simple and fast prediction methods like performance- predictive equations and closed-form solutions based on *Mechanistic-Empirical Pavement Design Guide* (MEPDG) simulations (Scott et al. 2013). However, the accuracy of these performance prediction methods was limited. As a result, agencies were hard pressed to defend the pay factors used to account for substandard pavement performance. The WesTrack project PRSs, although reasonable, introduced many unknown factors that may not have explicitly related to the quality of the product delivered by the contractor for a particular project. This shortcoming led other researchers to adopt other techniques. Since 2010, Applied Research Associates (ARA) and North Carolina State University under the sponsorship of Federal Highway Administration (FHWA) have tried to develop a PRS for hot mix asphalt using fundamental mechanistic models (Jeong et al. 2020). Material models based on viscoelastic theories to predict fatigue failure and permanent deformation were developed, a structure performance simulation software, i.e., FlexPAVE™, was released. The proposed methodology includes a performance-engineered mix design, a pavement design software, and performance- related specification. Under the methodology, mix design and QA were integrated for the first time. However, the mechanistic models and the performance tests were too complicated and time-consuming to be implemented by practitioners. Thus, it is critical to develop an implementable holistic methodology for improving asphalt mix durability in all three aspects: lab mix design, plant production QC/QA, and field placement.

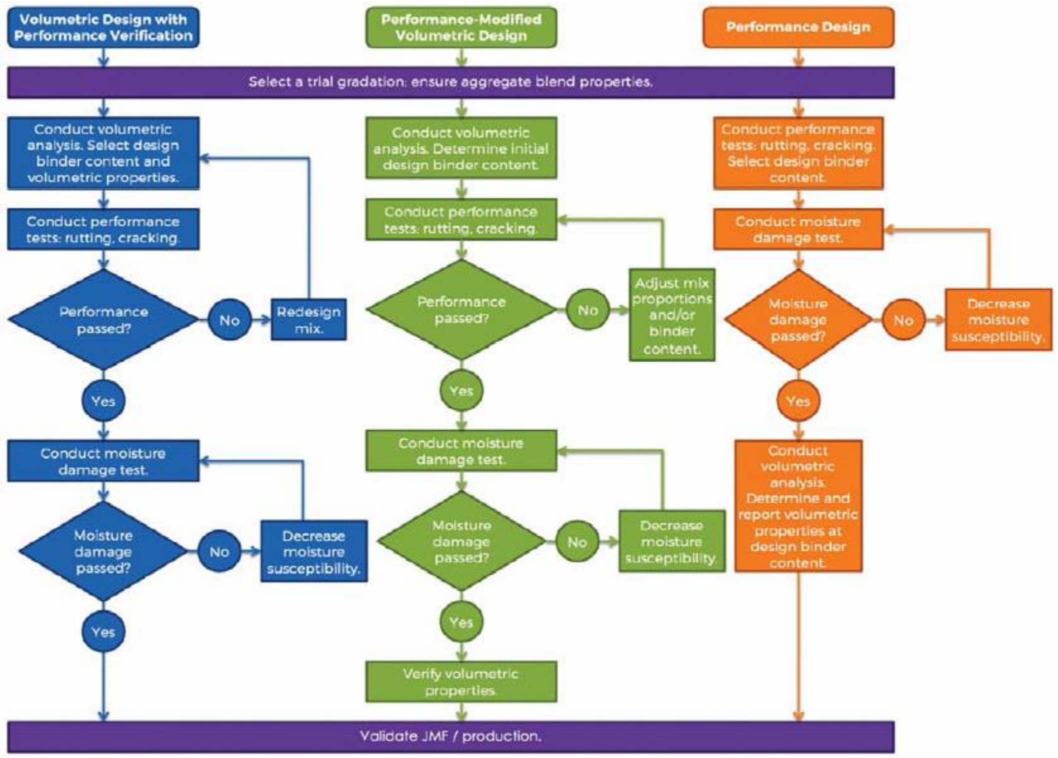


Figure 1.2 Three Implementation Approaches for the BMD Method (West et al. 2018)

1.2 Objectives

The goal and objective of this project is to develop a holistic methodology for improving asphalt mix durability in the areas of lab mix design, plant production, and field placement**.** This is directly tied to Thrust Area Number 4 of the research activities of the National TriDurLE: *Addressing aging and other materials-related distress of transportation infrastructure through the use of new materials, technologies, and construction methodologies*.

1.3 Methodology

To achieve the project goal and objective, the work was completed in several steps. A comprehensive literature review regarding the research and practice in developing performance mix design and performance construction specifications for QC/QA was first conducted. Both the technologies in terms of laboratory tests characterizing asphalt mix performance and approaches used in the lab and field for improving asphalt mix durability were reviewed and documented. At the end of this task, a state-of-the-art literature review paper was drafted and submitted to the *Journal of Transportation engineering, Part B: Pavement Engineering*. A draft framework of the proposed holistic methodology for balanced mix design and QA was then constructed. The framework aimed to provide potential methods to resolve the following problems:

- Defining critical distresses and selecting performance tests
- Developing rational performance test (or specification) requirements considering mix types and different applications
- Determining the role of volumetric properties in the BMD framework
- Aging or conditioning protocols for loose mixes and/or compacted samples
- Guidelines for selecting a balanced binder content
- Strategies for meeting performance requirements

The coherent BMD and QC/QA framework has four inter-connected components:

  - Volumetric mix design and selection of multiple asphalt contents for mixture performance evaluation,
  - Mixture performance evaluation at multiple asphalt contents and selection of the balanced asphalt content,
  - Mixture performance verification at the balanced asphalt content, and
  - Production QC/QA testing.

The research study also incorporated research efforts focusing on two important problems during asphalt mixture design and production: the testing variability within one laboratory and among different laboratories and the aging impact of the material performance. The research findings have been documented in this report.

1.4 Organization of the Report

The report documents the details of the research efforts and summarizes the findings in the study. Chapter 1 presents the objective and the methodology of the study, followed by a comprehensive literature review presented in Chapter 2. Chapter 3 includes the proposed holistic methodology of the framework for BMD and QA/QC with the recommendation of testing methods and aging conditioning procedure and a detailed case study. Chapter 4 presents study related to the variabilities of the rapid performance tests among different laboratories, and Chapter 5 compares the testing results and mixture performance at different aging levels during construction. The conclusions and recommendations can be found in Chapter 6.

Chapter 2 Literature Review

Pavement engineers have been seeking methods to improve the durability of asphalt mixtures ever since asphalt pavements were introduced. In the U.S., 4.1 million miles of public roads have been paved over the past few decades, and 350 million tons of asphalt mixtures are produced each year (NAPA 2019). Paving with durable materials can prevent premature pavement distresses, improve the road rideability and traffic safety, and benefit the taxpayer economically.

One of the early research efforts was the AASHO Test Road in the 1960s. The findings were incorporated in the 1972 Interim Design Guide and the 1993 AASHTO Guide for Design of Pavement Structures, which are still used by some state highway agencies (SHAs) as the primary pavement design guide. In the late 1980s, Congress approved $150 million to initiate the Strategic Highway Research Program (SHRP). In addition to the asphalt binder performance grading system developed as part of this work, an additional product of the research effort was the Superpave mix design method. It originally envisioned three hierarchical levels: volumetric-based mix design for mixtures targeting low traffic volume (Level 1) and volumetric and performance mix design (Levels 2 and 3) for mixtures targeting over one million Equivalent Standard Axial Loads (ESALs) of traffic.

However, due to the limited modeling and testing techniques at that time, only the Level 1 volumetric design method was implemented for all the traffic volumes. However, the efforts to develop "performance-based" asphalt pavement and mixture design methods continued. In the early 2000, the Mechanistic-Empirical Pavement Design Guide (MEPDG) was introduced. The design guide intended to directly use the predicted pavement performance from the mechanistic-empirical (ME) models based on the project local traffic and climate information to adjust the designed pavement structure. The MEPDG (now known as the AASHTO Pavement ME Design or the PMED) and other similar ME design methods were later incorporated in quality assurance (QA) methods and construction specifications. In recent years, as various performance testing methods have been introduced to the pavement industry, there has been a growing trend of including simple performance tests to design asphalt mixtures. Those efforts spawned the concepts of the balanced mix design (BMD).

As research studies on the above-mentioned topics have been conducted, remarkable review studies have been completed and acknowledged (West et al. 2018, Yin and West 2021, McCarthy et al. 2016). In this state-of-the-art review, the authors aim to organize and present the latest information related to the great efforts in incorporating performance in mix mixture design and mixture production, and most importantly, a new perspective with the attempt to integrate laboratory mix design, plant production, and field place acceptance is applied in the review. The challenges and solutions in the implementations are also summarized in the article.

2.1 Improving mixture durability under the framework of volumetric design

2.1.1 Past and current practices in Mix design Methods

Pavement engineers have been trying to understand asphalt mixture since the late 19th century. The early work focused on the development and usage of the volumetric parameters as well as some pass/fail tests. Some of these concepts were inherited by today's mix design methods. In the 1890s, the first mixture design formula and construction specification were published by F. V. Greene. In 1905, the concept of Voids in Mineral Aggregate (VMA), an important volumetric parameter that is still used in today's mix design specifications, was introduced by Clifford Richard, as well as pat-paper test' to determine the mixture gradation and binder content. Based upon the Richardson's design method, in the 1920s, the Hubbard Field design method was introduced, where the binder content was adjusted based on air voids, VMA, and mix stability. The philosophy of the design method was to have sufficient asphalt binder to satisfy the binder absorption in aggregates and form a binder film with a minimum thickness. The design used a test with motor oil to estimate binder absorption of aggregates and applied the 'Hveem stabilometer' to test the mixture stability. Compared to other mix design approaches (for example, the Marshall design), this method tended to yield lower binder content in mixtures and fewer rutting problems. The method is still used in some western U.S. states. The Marshall mix design method was developed in the late 1930s and early 1940s and adopted by the Corps of Engineers. The compaction method and the stability tests from the Hveem design were upgraded with automated devices. This method is currently widely used outside of the U.S., and its philosophy has been inherited by the Superpave Mix design method. (McDaniel et al. 2011)

In 1987, to deliver more durable mixes, the Strategic Highway Research Program (SHRP) was initiated. Fifty million dollars was allocated, which made SHRP the largest and most highly focused pavement research effort since the AASHO Road Test. As one major product of the SHRP program, the performance-based binder specification (based on rheology tests) was successfully implemented. However, on the mixture side, the Superpave mix design was finalized primarily based on volumetric parameters, which was originally planned for mixtures with design traffic volume of less than one million ESALs. The approaches drafted for higher traffic levels involved performance tests which were not be widely adopted due to the limitations in the modeling and testing technologies in the 1990s. The Superpave mixtures have been shown to perform better than previous mixes, and a benefit-cost study conducted in 1996-1997 quantitively showed that improved pavement performance and the increased pavement service life introduced by the new binder specification and the Superpave design method could bring a direct saving of $637 million per year or over 1.7 billion in 20 years if counting the reduced maintenance-related delays and vehicle maintenance (McDaniel et al. 2011). However, after years of practices, concerns on the Superpave volumetric design method have also been raised by pavement engineers. For example, it was reported that Superpave mixes often lack sufficient binder content for adequate durability (Maupin 2003) and the Superpave design is lack of proof tests to ensure the mixture performance (AAT 2011).

2.1.2 Improving Mixture durability by adjusting volumetric parameter

Over decades of usage, revisions to the volumetric parameters have been tried to produce more durable mixtures. VMA is one of the most important control parameters. A VMA within the proper range can provide sufficient space for effective binder and air voids as well as potential for adequate permeability (Asphalt Institute 2015). Some states have proposed to limit the range by having the maximum VMA values 1.5% to 2.0% above the minimum values and remove the upper limit of VFA to simplify the design procedure (Christensen, Jr. and Bonaquist 2006). While attempting to adjust the VMA limit, one needs to consider the interactions among the volumetric parameters. For example, adjusting the aggregate structure to target a higher VMA may decrease the mixture compatibility; thus, potentially leading to lack of compaction or a low density in the field (Christensen, Jr. and Bonaquist 2006). Voids filled with asphalt (VFA) and Effective binder content (Vbe) are also important factors. Some states attempted to adjust the minimum requirements for VFA to ensure sufficient binder in the mixture (AAT 2001). Alternatively, some engineers proposed to use the concept of binder film thickness to determine the optimum binder content. While some researchers believed that there was no physically existing film in compacted mixtures, others thought they did exist in loose mixtures and the calculated ‘apparent film thickness’ had higher correlations with mixture performance than the conventional VFA. Researchers found that apparent film thickness between 7 to 9 microns could yield both suitable workability and rut resistance (AAT 2011). In the National Cooperative Highway Research Program (NCHRP) 9-25 Project, the fineness modulus

(FM300) was introduced to quantitatively represent the effective surface area of aggregates. It was believed that the effective surface area can be used to calculate the film thickness and further to determine the binder content (Newcomb et al. 2015).

Air voids at the design number of gyrations (Ndes) or the design air voids have also been adjusted. The design air void was fixed at 4% in Superpave mix design. The Indiana Department of Transportation (INDOT) showed that increasing the design air void to 5% can increase the average mixture density compacted in the field from 93.3% to 95.3% (Montoya et al. 2018). Broadening the design air void content from a fixed value of 4.0% to a range of 3.0% to 5.0% is another common adjustment among SHAs (Christensen, Jr. and Bonaquist 2006). The associated additional compaction effort and stability are believed to be the major benefits of increasing the design air voids in most of cases. However, it is important to know that, like VMA, the design air void is a function of many other factors, i.e., the gradation, binder content, and the compaction energy. Some researchers pointed out that increasing the design air void might lead to an aggregate structure that is hard to compact as well as lower effective binder content, which might compromise the expected benefits. (AAT 2011).

In addition to the volumetric variables in asphalt mixtures, engineers have also attempted to adjust other parameters, for example, Ndes. Some studies suggested increasing the Ndes by one level to request more compaction energy in the field (Christensen and Banaquist 2006). Meanwhile, others, including the Utah DOT, proposed to decrease the Ndes to increase the design binder content (AAT 2011, Tran et al. 2016). Essentially, the mechanism and the results of adjusting the Ndes is similar to changing the design air voids. Increasing or lowering the required values may lead to the expected improved mixes, but the adjustments should be determined carefully with the consideration of the changes and interactions of other volumetric parameters.

The dust-to-binder ratio or the dust ratio is specified with a limit of 0.6 – 1.2 in the current Superpave design. This parameter is related to the specific aggregate surface area. A survey conducted by the New Jersey DOT showed that most of the states have adjusted the limit in their design specifications (NJDOT 2011). The NCHRP 9-25 Project mentioned that some adjustments in the dust-to-binder should be able to reduce the mixture permeability. In a study conducted for the Colorado DOT, an increase in dust-to-binder ratio was suggested to account for the 1 percent hydrated lime (Scott 2019).

2.1.3 Quality assurance under the framework of volumetric design

According to a survey by McCarthy et al. (2016), quality assurance (QA) is the most popular type of construction specification. In this system, mix design, mix acceptance, and QA usually share the same set of volumetric-based parameters and have consistent threshold limits. The pay adjustments or the incentives/disincentives are usually determined based on the percentage within limits (PWL). However, since the early 2000s, almost one decade after the Superpave was deployed, the pavement community has realized that using volumetric parameters only is not sufficient to ensure mixture durability and performance, especially when recycled materials and innovative composites are involved. One solution is to incorporate the mixture performance directly in mix design and construction specification. The efforts to develop performance mix design and performance specification are documented in the next section.

2.2 Development of performance specifications

A construction specification includes information and methodologies to complete the project (i.e., project initiation, bidding, design, production and placement, and acceptance and payments) (TRB 2005, 2018). The developments and the merits of different types of pavement construction specifications, including the performance specifications, are briefly documented in this section.

2.2.1 Construction specification types and methods

Figure 2.1 lists existing types of construction specifications. From left to right, the risk in a construction project gradually shifts from the agency to the contractor. The earliest specification was the method specification (or the “recipe specification), which required the contractor to produce and place a product using the specified materials in definite proportions and specific types of equipment and methods under the direction of the agency. It could not incorporate the construction variability and did not allow any innovations from the contractors. In the 1950s, the End-Result Specifications were introduced along with the construction of the AASHO Road Test. It required the contractor to take the entire responsibility for producing and placing a product. The agency would either accept or reject the final product or apply a price adjustment commensurate with the degree of compliance with the specification. Although it provided the contractor some flexibility, the agency would take the risk of having to reject a large quantity of materials at the end of the project.

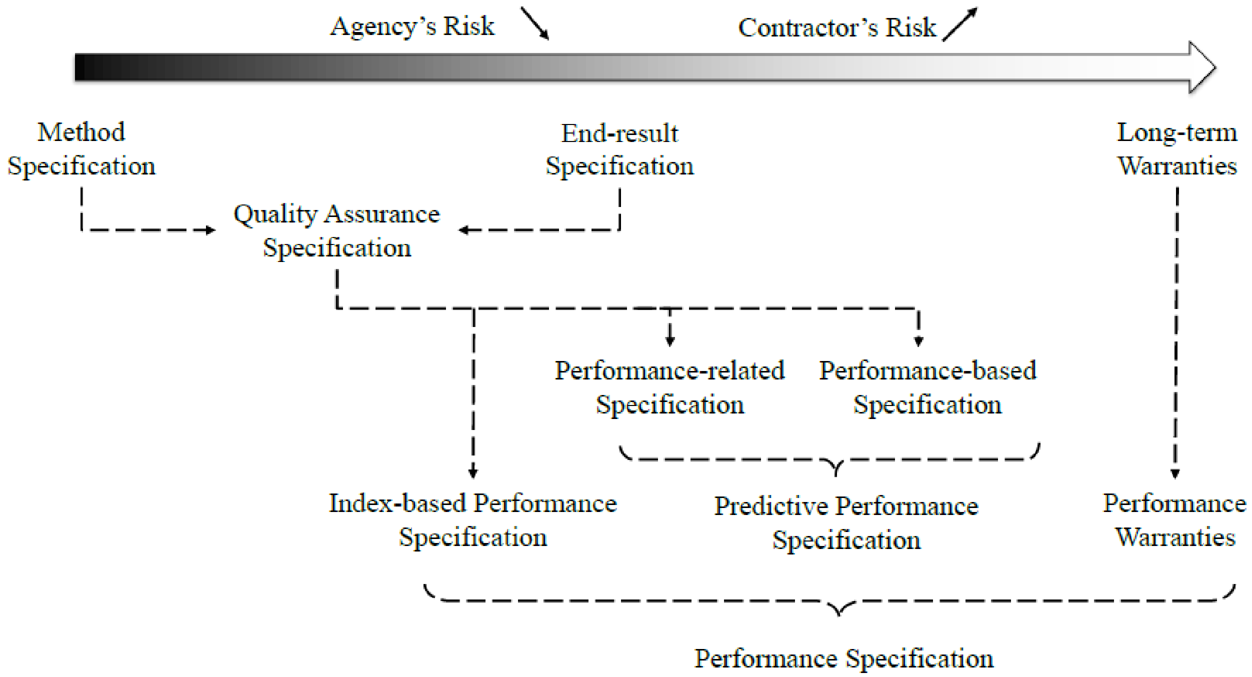


Figure 2.1 Relationships among different types of construction specifications

Since its introduction in the 1960s, the QA specifications have been prevalent among SHAs. According to a Ksaibati and Butts (2003), among the 45 SHAs that responded, 40 U.S. state agencies had adopted QA specifications. It has inherited the merits of both method specifications and end-result specifications by requiring the contractor to conduct Quality Control (QC) and agency to perform acceptance activities throughout production and placement of a product. The final acceptance of the product is usually based on a statistical sampling of the measured quality level for key Acceptance Quality Characteristics (AQCs). While the contractors are granted the flexibility for innovations in their products, the agencies have the control of the key AQCs and the product quality on real-time basis during the production. Another type of specifications is the warranty specifications, which guarantees the integrity of a product and assigns responsibilities for the repair or replacement of defects to the contractor (TRB 2018).

In the 1990s, the component of pavement performance was proposed to be added to the QA specifications, which led to the development of development the Performance-Related Specifications (PRS) and Performance-Based Specifications (PBS). PBS is defined as "a QA specification that describes the desired levels of fundamental engineering properties that are predictors of performance and appear in primary prediction relationships (i.e., models that can be used to predict stress, distress, or performance from combinations of predictors that represent traffic, environment supporting materials, and structural conditions)" (TRB 2005, TRB 2018). The fundamental engineering properties include but are not limited to dynamic modulus, creep properties, and fatigue properties. In a PBS, the acceptance should be based on the measurement of the fundamental engineering properties of the finished product instead of the AQCs that are indirectly related to performance, and the pay adjustment should be determined based on the difference between the as-designed Life-Cycle Cost (LCC) and the as-built or as-constructed LCC. However, no true PBS for hot mix asphalt (HMA) has been implemented because most of the fundamental engineering properties cannot be measured within the time requirements during production (TRB 2018). Given the challenges in developing PBS, PRS is an alternative solution.

PRS is defined as "a QA specification that uses quantified quality characteristics and LCC relationship that are correlated to product performance" (AASHTO 2003, TRB 2005). In other words, PRS can use AQCs to estimate or predict the fundamental engineering properties. Like the PBSs, pay factors in PRSs are determined based on pavement life predicted using the estimated fundamental properties. Along with the performance warranty specifications, which are derived from the long-term warranty specifications, the PBSs and PRSs form the performance specification family.

As simple performance tests become prevalent, pavement engineers have been trying to incorporate these tests into QA specifications. Instead of predicting pavement performance using mechanistic models and ME programs, the performance tests use index parameters to evaluate mixture performance and durability. The relevant QA specifications are also written based on the performance indices. In this report, those type of QA specifications are designated as index-based performance specifications (IPS). Some researchers have named their index-based specifications as PBS; to avoid confusion, in this article, they are referred as IPSs. The integration of mix design and performance specification will be discussed later in this report.

2.2.2 Development of performance specifications

2.2.2.1 Predictive performance specifications

Predictive performance specifications, including the PRS and PBS for HMAs, are the QA specifications that base the acceptance and the pay adjustments on the predicted pavement performance. Starting from the early 2000s, the research PRS and PBS intended to utilize the Mechanistic-Empirical (M-E) performance predictive models. A PRS framework was developed using the WesTrack test road data during the NCHRP 9-20 project. One major product was the software package, HMA Spec. It could generate the project-specific PRS and pay adjustments based on the differences between the as-design and the as-constructed pavement performance. The program used the stiffness, permanent deformation, and fatigue cracking as the three primary variables to determine the pay factors. The performance predictive models considered factors including the construction material, environment, traffic, and roadbed soil. The cracking and rutting model were developed based on the regression analysis using the collected field data (Epps et al. 2002).

The NCHRP 09-22 project was later launched to conduct further investigation on the PRS framework (Fugro Consultants 2011). The main difference between the proposed PRS in the new research project and the previous one was the predictive performance models. The MEPDG program (ARA 2004) was adopted to predict the as-design and the as-constructed pavement lives. A PRS program called the Quality Related Specification Software (QRSS) was developed to generate project specific PRSs. The predicted pavement distresses i.e., rutting, fatigue cracking, thermal cracking, and rideability quantified by the International Roughness Index (IRI) were used to adjust the payments in the specification. The AQCs that were required to estimate fundamental engineering properties included air voids, asphalt content, aggregate gradation, volumetric properties, and the binder viscosity of the AC layer. The program used the probabilistic method considering the variabilities in the construction and laboratory measurements to the pay factors. The method applied the Monte-Carlo simulations for rutting and fatigue cracking while the Rosenblueth probabilistic point estimate method was used for thermal cracking analysis. The pay factors were determined based on the PF-PLD relationship (where PLD stands for predicted life differences). The final incentive and disincentive were calculated in dollars by considering all individual pay factors.

Due to the limitations in the prediction accuracy of the MEPDG models and the efforts required to calibrate the models, from 2008 to 2021, the Federal Highway Administration (FHWA) funded Applied Research Associates, Inc. (ARA) and North Carolina State University (NCSU) to develop an improved PRS framework (Kim et al. 2017, Kim et al. 2021). The outcomes of the research included the development of three fundamental material models and the corresponding material testing protocols, a structural performance simulation program, a performance mix design framework, and a QA strategy. Among the three material models, both the fatigue cracking model, i.e., the Simplified ViscoElastic Continuum Damage (S-VECD) fatigue model, and the low-temperature cracking model stemmed from the VECD theory, and one cyclic fatigue test was able to calibrate the model coefficients for both models.

The models used fundamental material properties, such as the describe the relationship between the reduction of material pseudo stiffness (C) and the growth of damage (S) (Underwood et al. 2010, Wang and Kim 2017, Ashouri et al. 2021), which were independent of loading conditions (i.e., modes of loading, loading amplitudes, frequency, and loading temperatures) and could be used to predict material behaviors under different circumstances. The rutting model was related to the permanent strain formed under different loading amplitudes with virous resting periods at different temperatures. The materials models were implemented in the structural performance simulation program, FlexPAVE™, which conducted three-dimensional viscoelastic analysis with the consideration of the in-situ pavement structure, traffic, and climate data. Good agreements have been found between the FlexPAVE™ predictions and field measurements (Wang et al. 2021b, Wang et al. 2016, Wang et al. 2018). To determine incentive/disincentive based on predicted performance and save the material testing time during QA, a bridge connecting the routine AQC and the predicted performance, i.e., the performance-volumetric relationship (PVR) was developed (Wang et al. 2019). During production, the contractor and the agency only needed to measure the volumetric properties, same as in the existing QA procedures, and the PLD for each lot could be predicted using PVR. Shadow projects using the PRS framework have been deployed in some U.S. states (Jeong et al. 2020, Kim et al. 2021); however, extensive efforts for training and demonstration would still be need for future implementation.

In addition to the research efforts supported by the NCHRP and FHWA, other researchers have also developed performance specifications that were suitable for local applications. Since the late 1990s, California has been spending efforts to develop PBS based on the M-E methods. The original performance mix design and construction specification incorporated performance tests, i.e., the flexural beam test and the repeated simple shear test. In 2000, the CalME flexible pavement design software was first introduced. It was developed based on incremental-recursive damage models regarding the fatigue and rutting performance (Harvey et al. 2014). The CalME program calculated the pavement fatigue life with the given material and structural conditions. The reliability of the predictions was evaluated using the Monte-Carlo analysis. The CalME calculated the mechanical responses using the linear layered-elastic-based ELSYM5 program. The obtained largest maximum principal tensile strain was then used to evaluate the temperature equivalency factor (TEF) and temperature conversion factor (TCF). The pavement fatigue life was predicted using the three-stage Weibull equation (Tsai et al. 2012) with the reliability and the effects of pavement structure and climate taken account.

In 2004, Williams et al. (2004) developed a PRS for the Michigan DOT. The specification used predictive equations and relationships between field performance (i.e., rutting and cracking) and mix properties (i.e., air voids and binder content) to determine the pay factors. In the proposed PRS framework, the testing methods included the four-point beam fatigue test and the asphalt pavement analyzer test, and the obtained performance indices were correlated with field performance based on the past testing experience. The sampling methods were thoroughly discussed in the study, and the existing mix designs were verified.

In summary, many researchers and organizations have participated to establish PRS framework in the past two decades. The proposed PRSs utilize either mechanistic or empirical material and structural models to predict pavement lives. These approaches have the merits of both the performance-based specifications and the conventional QA specifications. On the other hand, the challenges to deploy the PRSs include to effectively quantify and encompass the risk and reliability in the specifications (Hughes 2005, Hughes et al. 2012), to provide training and instructions for SHAs and contractors, to develop local material database and local model calibration coefficients, and to simplify the performance tests. Some of these problems have been mitigated in the index-based performance specifications.

2.2.2.2 Index-based performance specifications (IPS)

Since the late 2000s, various simple mixture performance tests have been introduced to the asphalt community. The tests usually define indices to evaluate the mixture durability or its resistance to a certain type of pavement distress, and most of the tests can be easily and quickly performed compared to the mechanistic-based material characterization tests. The index-based performance specifications or the IPSs have been introduced with the integration of the simple performance tests and the QA specifications.

In 2016, Mohammod et al. (2016b) developed a standard PBS method and a simplified PBS approach. The standard PBS was based on the AASHTO Pavement ME Design Software and considered the pavement structure, climate, traffic, and material properties (i.e., dynamic modulus). The simplified PBS was an IPS which required the rutting and cracking test results. The comparison between the standard PBS approach and the simplified PBS approach suggested the simplified PBS was recommended because the Pavement ME program had not been locally calibrated.

Given the limitations in the volumetric-based mix design and QA specifications and to accommodate the usage of Reclaimed Asphalt Pavement (RAP), the New Jersey DOT (NJDOT) started to develop performance specifications with performance-based acceptance procedures (Bennert et al. 2014). In the proposed IPS, the contractor should design mixtures passing the volumetric criteria, and the agency conducts performance tests with the submitted component materials. Once the mixture is approved, the contractor should construct a test strip where plant-produced mixtures are sampled and tested. If the mixture meets the performance criteria, the production can be continued, and the mixtures should be sampled and tested at a certain frequency during production. The Asphalt Pavement Analyzer (APA) should be used to assess the mixture rutting resistance, and either the Flexural Beam Fatigue test or the Overlay test should be adopted for the cracking resistance evaluation depending on the mixture type. The NJDOT has successfully developed five performance-based asphalt mixtures, namely, high-performance thin overlay (HPTO), binder-rich intermediate course (BRIC), bridge deck waterproofing surface course (BDWSC), bottom-rich base course (BRBC), and high RAP (HRAP) (Bennert et al. 2014).

The Northeast Pavement Preservation Partnership (NEPPP) developed a pilot specification for HPTO mixtures. The specification encompassed surface preparation, material properties, mixture design requirements, RAP testing requirements, and mixture performance criteria. Performance tests, i.e., the Overlay test, the thermal stress restrained specimen test (TSRST), the four-point flexural beam fatigue test, and the APA test, were required to be conducted on plant-mixed mixture during production and placement. The mixtures must meet both the performance index limits and the Superpave volumetric requirements. After two years of field monitoring, the HPTO mixture from New Hampshire performed well with minimal cracking observed; in contrast, the conventional mixture had 25% cracking returned. No apparent distress was found in the Vermont test sections for both HPTO mixtures with and without RAP (Mogawer et al. 2012).

In summary, the IPSs are QA specifications with simple performance tests incorporated. Compared to the predictive performance specifications, the IPSs are more intuitive and can adopt performance tests that can be completed in a timely manner. However, there are still challenges in implementing the IPSs, including the relatively high cost and uncertainties, lack of communication between the SHA and the contractors, and the delay in production for conducting performance tests (McCarthy et al. 2016).

2.3 Development of performance mix design methods

2.3.1 Performance mix design approaches

In the 2000s, several years after the Superpave mix design was widely implemented in the U.S., multiple simple performance tests were developed and incorporated into mix design methods, including the prevalent balanced mix design (BMD). Like the performance specifications, in this article, the performance mix design methods are introduced as index-based performance mix design (IPMD) and predictive performance design. The existing BMD methods are IPMDs as the incorporated performance testing methods are index or tolerance tests.

2.3.1.1 Index-based performance mix design/balanced mix design

The concept of BMD was first proposed in 2007, as a ‘balanced binder content’ was expected to be identified for given mix components through index-based performance tests so that the designed mix was neither too ‘lean’ to form pavement fatigue cracking or too ‘wet’ to yield deep rut depth (Zhou et al. 2007). In September 2015, the FHWA Expert Task Group (ETD) on Mixtures and Construction founded a Balanced Mix Design (BMD) Task Force. The BMD task force defined BMD as “asphalt mix design using performance tests on appropriately conditioned specimens that address multiple modes of distress taking into consideration mix aging, traffic climate and location within the pavement structure” (West et al. 2018). Three pathways were originally developed by the task force, and later, they were expanded to four approaches when the corresponding provisional AASHTO standards PP105-20 and MP 46-20 were submitted, as presented in Figure 2.2 (Yin and West 2021). The details and the highlights in each approach are demonstrated in Table 2.1. Under the BMD framework, several implementation plans have been developed among SHAs (Paye 2014, Cross and Li 2019, ALDOT 2020, Bennert 2020, Coleri et al. 2020).

Like the IPSs, the advantages of IPMD or BMD include the relatively short turnaround time of the simple performance tests and the intuitive design philosophy. Using both the volumetric and performance criteria provides the pavement engineers confidence of the mixture quality. However, the index threshold limits can only be determined based on empirical relationships between field performance and performance test results. Besides, the indices cannot take the project-specific information (i.e., structural, environment, and traffic conditions) into account, and neither can the formation of pavement distress as a function of service time be predicted using the index-based approaches, as envisioned in the original SHRP project.

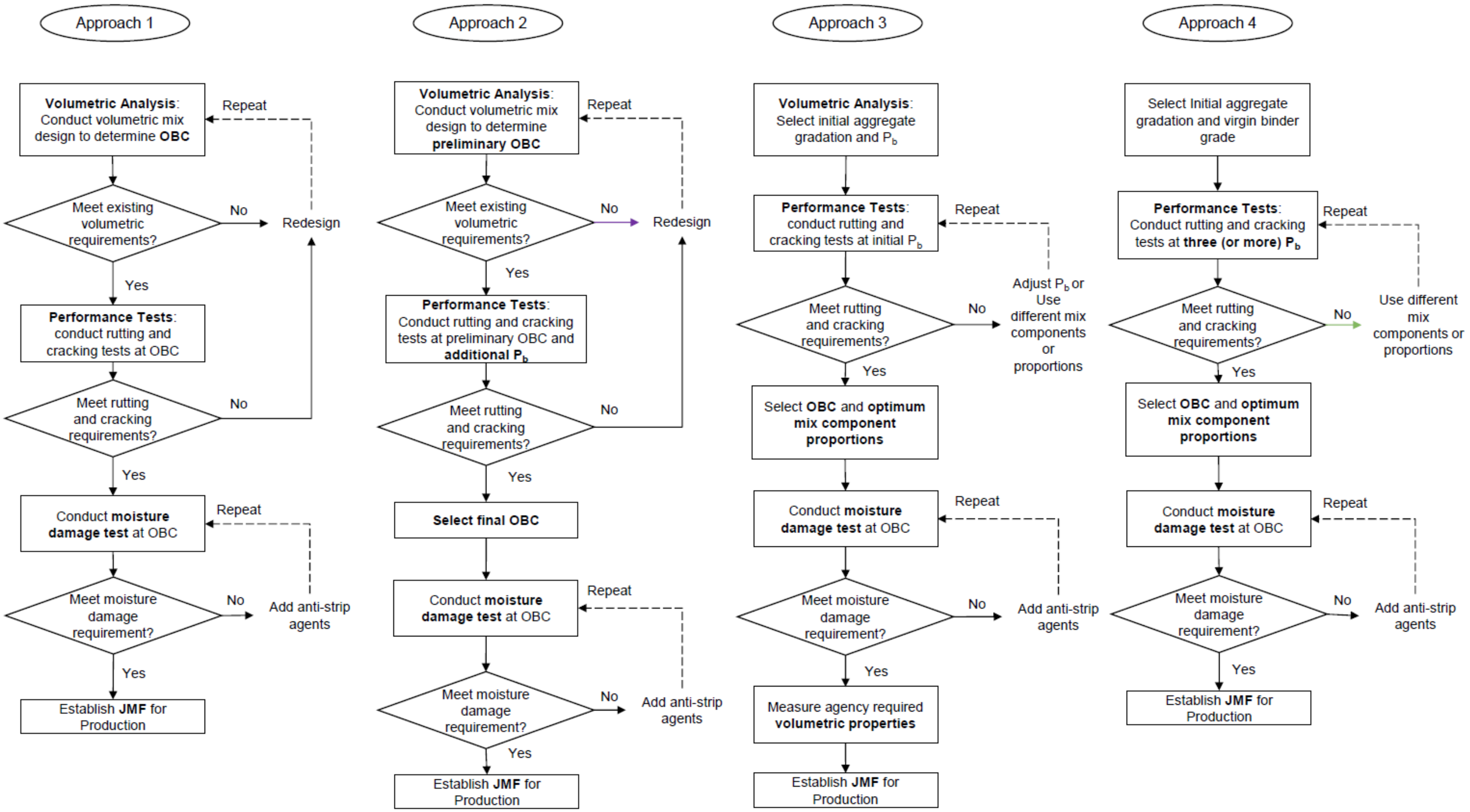


Figure 2.2 Flowcharts demonstrating BMD approaches (Yin and West 2021)

Table 2.1 Highlights of Different BMD Approaches

Approach	Description	Highlights
BMD Approach 1	Volumetric Design with Performance Verification Approach	• This approach makes sure that all the mix designs are products of volumetric design methods with performance requirements satisfied. • This approach applies additional constraints for performance requirements onto the original volumetric designs. This combination provides the engineers the most confidence but least design flexibility for the contractors.
BMD Approach 2	Volumetric Design with Performance Optimization Approach	• This approach is an expanded version of Approach 1, and it was not included in the original three approaches proposed by the former FHWA BMD Task Force. • This approach allows a potential offset in optimum binder content determined based on the performance test results from volumetric optimum binder content while the mixture gradation and other mix components will remain the same as designed by the volumetric-based method. • When this approach is adopted, the binder contents for performance testing will usually be preliminary OBC - 0.5%, preliminary OBC, preliminary OBC + 0.5%, and preliminary OBC + 1.0%.
BMD Approach 3	Performance-Modified Volumetric Design Approach	• This approach and Approach 1 both start with volumetric design. • Unlike Approach 1, this approach allows adjustments for both the gradation and binder content based on the performance test results. The final combination of the gradation and binder content is not directly obtained from volumetric design, and only some volumetric criteria are required to be met.
BMD Approach 4	Performance Design Approach	• This approach is similar to the third approach proposed by the former FHWA BMD Task Force, but more details and instructions were provided than the descriptions when it was first introduced by the task force. • This approach is a combination of Approach 2 and Approach 3, and it may not necessarily start with a volumetric design. • After the initial selection of an initial selection of aggregate gradation, recycled asphalt materials, content, and virgin binder grade is determined, the binder contents for performance testing will usually be initial binder content - 0.5%, initial binder content, initial binder content + 0.5%, and initial binder content + 1.0%.

2.3.1.2 Predictive performance mix design

Predictive mix design utilizes the mixture/pavement performance predicted from the mechanistic models to determine the optimum mixture design. One predictive design approach has been proposed by NCSU. The method is also known as the Performance-Engineered Mix Design (PEMD) (Wang 2019, Kim et al. 2021, Wang et al. 2021a). One feature of the method is that instead of using the trial-and-error approach (creating one trial design and using volumetric or performance criteria to determine pass or fail), the PEMD identifies the performance-optimum design directly from the infinite numbers of combinations of the given material components. The identification of the optimum design can be achieved by using the PVR. PVR characterizes a given mix design with two variables, the in-place VMA (VMA_{IP}) and in-place VFA (VFA_{IP}) and forms a two-dimensional volumetric space. Each point in the space is corresponding to one combination of gradation, binder content, and compaction level, as presented in Figure 2.3**Error! Reference source not found.** (a). Previous research (Wang et al. 2019) found that when the climate, traffic, and structural conditions are known, the predicted pavement performance (% fatigue damage and permanent deformation) is a bilinear function of VMA_{IP} and VFA_{IP}. The PVR function, therefore, provides a spectrum of performance as a function of the change of gradation, binder content, and compaction level, as presented in Figure 2.3**Error! Reference source not found. Error! Reference source not found.** (b). The performance optimum design can be acquired by combining the predicted pavement life determined by fatigue life and rutting failure (Wang 2019, Wang et al. 2021a). It will be the SHA's discretion to require all or some of the volumetric limits to be met, and the moisture susceptibility of the design mixture can be tested afterward. The design procedure requires four sets of performance tests to calibrate the PVR function coefficients. The performance tests used in PEMD are the cyclic fatigue test and the SSR test. If the design candidate fails the volumetric or the moisture tests, the designer can select another combination from the volumetric spectrum without conducting additional performance tests. Therefore, unlike the unknown numbers of iterations that the IPMD methods may require, with the fixed number of performance tests and testing time, the design timeline can be planned by the SHAs and contractors. Twelve days are expected to complete the design including the specimen preparation and testing time (Wang et al. 2021).

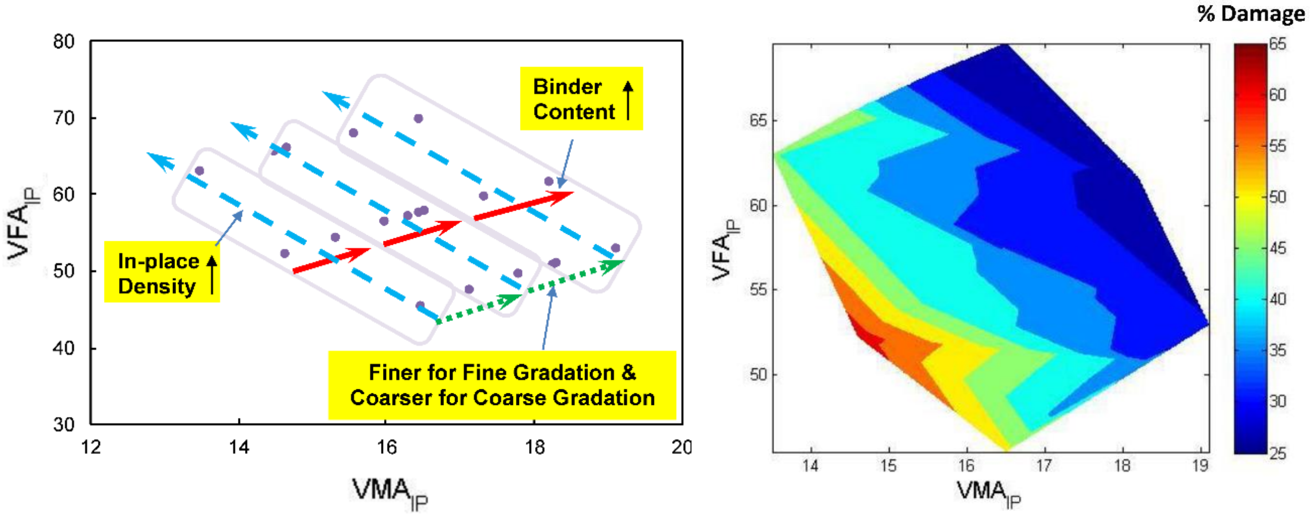


Figure 2.3 (a) Volumetric space formed by VMA_{IP} and VFA_{IP} and effects of mix design parameters on VMA_{IP} and VFA_{IP} and (b) %Damage contour in a pavement structure in the volumetric space (Wang et al. 2019)

2.3.2 Comparison of performance Mix design approaches

Table 2.2 summarized several important criteria to evaluate the performance design approach. Among the four BMD approaches, the involvement of the performance increases from Approach 1 to Approach 4 as the volumetric restrictions for binder contents and gradations are gradually released. As a result, Approach 1 provides the agency highest confidence and lowest risks when BMD is first implemented. However, Approach 1 meanwhile grants the contractor the least design flexibility. As for the predicative performance mix design methods, the performance involvements and design flexibility would be the highest among all the methods; however, the agency may have the least confidence since the design will primarily be based on the predicted performance instead of the conventional volumetric parameters.

In terms of the design effort needed in each approach, the predictive performance mix design method may consume the longest time because of complexity of the required performance tests.

However, the BMD approaches may take longer time if multiple iterations are needed for redesign or adjustments; in contrast, the number of performance tests is fixed for the predictive performance design method. As for the development of performance specification, in each approach, the same performance requirements can be consistently applied in mix design and in QA as long as the specifications are well engineered.

Table 2.2 Factors to Consider for Comparing Performance Design Approaches

Evaluation Criteria	Description
Performance Involvement	The extent of performance involved in the design procedure. If the performance tests are only used for pass/fail decision after volumetric design, the approach will be evaluated with low performance involvement. High involvements are granted to the methods where gradation, binder content, and other design parameters are determined based on the mix performance.
Confidence and Risk	The confidence level that SHA has when they first switch from volumetric mix design to this approach. For example, if volumetric limits are not required in a new mix design, the SHA would have less confidence and take higher risk on the mix than on the mixtures that meet the volumetric requirements.
Design Flexibility	The flexibility that the contractor has while conducting the mix design. For example, if the volumetric requirements are inherited, the additional constraints added for performance will further limit the design flexibilities for contractors. Higher design flexibility can provide contractors incentives to apply innovative materials and technologies.
Design Effort	The design effort indicates the time and resources that the performance design approach costs. This criterion should be evaluated based on the number of gyratory specimens required to be fabricated, the number of performance tests, the testing turnaround time, and the estimated design time in days.
Performance Prediction Capacity	The performance prediction capacity indicates how well the mixture performance is incorporated in the mix design approach. Does the mix design rigorously consider the target pavement structure, climate, traffic volume, and other factors? Can the mix design approach predict the pavement distress deterioration with time as initially expected in the Superpave Level III design? Are the performance threshold limits determined by empirical or mechanistic method?
Compatibility with QA	The compatibility with QA index indicates the difficulty level to develop a Quality Assurance method and a performance construction specification with pay adjustments using the same testing methods and/or performance threshold limits.

2.3.3 Integrating performance mix design with QA

As both mix design and QA are important steps on the asphalt production chain, it is necessary to develop holistic methodologies to accommodate performance mix design and QA methods. Consistent evaluation strategies and criteria would be preferred for the system. As BMD methods are implemented, great efforts have been made by researchers and SHAs to form such coherent systems. For example, Zhou et al. (2020, 2021) proposed a framework. It started with a mix design method where the volumetric criteria are used to obtain the initial candidate and performance tests are conducted at multiple binder content levels to determine the final binder content. The agency was required to conduct performance tests on the submitted mix design for verification and acceptance. During production, the plant-mixed mixtures would be sampled for QA testing. The same performance tests and performance index limits as in the mix design were suggested for QA. To meet the time requirement, simple tests, such as the IDEAL-CT, IDEAL-RT, and HWTT, were proposed to assess the mixture cracking resistance, rutting resistance, and moisture susceptibility, respectively. Meanwhile, several states have specified performance requirements during production, as presented in Table 2.3.

Table 2.3 Summary of State-of-the-Practice on BMD Implementation (Yin and West 2021)

BMD Approach	State	Rutting Test	Cracking Test	Performance Testing for Production Acceptance?
Approach 1	Illinois	HWTT	I-FIT	Yes, HWTT for “Pass/Fail”
	Louisiana	HWTT	SCB-Jc	Yes, “Pass/Fail”
	New Jersey	APA	OT*, BBF	Yes, “Pass/Fail” or Pay Adjustment
	Texas	HWTT	OT, IDEAL-CT	Yes, “Pass/Fail”
	Vermont	HWTT	I-FIT	Yes, PWL
Approach 1 and 4	Virginia	APA	Cantabro, IDEAL-CT	Yes, “Pass/Fail”
Approach 3	California	FN, HWTT	BBF, I-FIT	Yes, HWTT for “Pass/Fail”
	Missouri	HWTT	I-FIT, IDEAL-CT	Yes, HWTT for “Pass/Fail”, I-FIT & IDEAL-CT for Pay Adjustment
	Oklahoma	HWTT	IDEAL-CT	No
Approach 4	Alabama	HT-IDT	AL-CT	Yes, “Pass/Fail”
	Tennessee	HWTT	IDEAL-CT	To be determined

As for the predictive performance designs and specifications which rely on performance testing results to predict mix performance and determine pay factors, the testing turnaround time during production would be the main challenge to develop the coherent framework. To integrate the PRS and PEMD, one strategy that was applied was to use the PVR (Jeong et al. 2020). While the volumetric parameters were measured as AQCs during production, with PVR, the performance of the mixes in each lot could also be estimated. To accommodate the variabilities in materials, the performance tests were suggested to be conducted once in every several thousands of tons to calibrate the PVR coefficients. The challenges for implementing and integrating the performance mix design and performance construction specifications are discussed in the following section.

2.4 Challenges and solution in development of holistic methods

Introducing the performance specifications/mix design methods will make changes to the process of bidding, mix design, acceptance, production, and payment/reward and penalty. Though some simple performance tests have been developed, one major challenge is still their turnaround time compared to the conventional volumetric tests. Also, depending on the type of the performance specification/mix design method that the SHA uses, some extent of communication and training are also necessary at the initial stage of the deployment. Selecting the type of performance specification and mix design method that is suitable for the local area is usually the SHA's responsibility. If an index-based mix design method will be applied, the SHA also needs to determine the design approach (Approaches 1, 2, 3, or 4), the performance testing methods, and the corresponding threshold values. Some other potential challenges in developing and deploying performance specification/mix designs are listed in the following (Harvey et al. 2014, McCarthy et al. 2016, Diffenderfer and Bowers 2019, Lee et al. 2020, Kim et al. 2021):

- Lack of experience and confidence in developing and using the new specification/design methods;
- Lack of historical data to develop material testing database, calibrate prediction models, and determine index threshold limits;
- Needs to quantitatively estimate of the risk and reliability in the specification and design method in a multi-layer pavement system;
- Needs to specify the details in the specifications/mix design methods, such as the sampling position and sampling frequency, type of mixture for performance testing (laboratory mix vs. plant mix), and aging conditions in performance tests (short-term aging vs. long-term aging);

and

- Needs to evaluate and account the effects of the testing variabilities in the specification/mix design.

The following discussion focuses on the details and solutions to some of the common challenges in developing performance specification/mix design.

2.4.1 Selection of performance testing methods

Selecting the performance testing methods is one major step to develop an IPS or IPMD. The performance tests should be corresponding to the major types of pavement distresses in the region (Diffenderfer and Bowers 2019, Yin and West 2021). There are multiple cracking and rutting tests that can be used in an IPS or IPMD. Some of the tests may require longer testing time but provide the material fundamental engineering properties along with performance indices, such as the Cyclic Fatigue test and the Disc-shaped compact tension (DCT) test. Some of the tests are tolerance tests performed under a single testing condition. A parameter obtained from the test is usually used as the index representing the material's resistance to a certain pavement distress. When selecting the testing methods, SHAs should consider key factors such as users' experience, availability of existing data, availability of testing equipment, cost of the test, turnaround time, and the effectiveness of the test. Table 2.4 presents some commonly used performance testing methods. West et al. (2018) provided a nine-step guide for the determining the testing methods in BMD. The features of each testing method have been documented in articles (McCarthy et al. 2016, Lee et al. 2020, Yin and West 2021) and AASHTO MP 46-20.

To develop a BMD protocol for Virginia, the state DOT considered the Cantabro test (for its uniqueness in testing durability and simplicity), the APA test (due to its historical application in the state), the Overlay test (to test resistance to reflective cracking), the I-FIT test (based on previous research results), the N_{flex} factor test (based on available research results in West et al. (2017) and its simplicity), and the IDEAL-CT test (for its simplicity) as candidate performance testing methods. During the evaluation, factors including the test effectiveness (by comparing test results with known mixture performance), cost in running each test, state-wide equipment distribution, training requirements, specimen preparation time, test repeatability were considered.

The Cantabro test, the APA test, and the IDEAL-CT tests were eventually selected (Diefenderfer and Bowers 2019, Diefenderfer et al. 2021).

Climate is another factor for the determination of performance testing methods. In Minnesota, the most common pavement distress was found to be low temperature cracking due to its cold weather.

Therefore, the Minnesota DOT included the DCT fracture energy test in specification provision for BMD (Johanneck et al. 2015). Likewise, in addition to the HWTT and the SCB test, the DCT test was also required the Wisconsin DOT (Paye 2014).

SHAs usually validate the performance testing results with their local mixtures. One method is to compare the ranking of the mixtures obtained from the testing results with the known mixture performance. Another approach is to conduct performance tests on local mixtures with systematically varying design parameters. The candidate test should be able to distinguish mixtures with different properties and design variables (such as different binder contents, air voids, gradation, RAP contents, aging levels, binder type, and so forth). In the research conducted by the Indiana DOT, the I-FIT SCB test and the HWTT were evaluated as the primary candidate testing methods for BMD (Lee et al. 2020). The effectiveness of the tests was assessed by their sensitivity to different design variables, which included the specimen air void, specimen geometry, and binder performance grade. The results indicated that the I-FIT did not follow the engineers' intuition that asphalt specimens with lower air void contents or modified binders should yield higher FI values. Therefore, the I-FIT test was determined not to be implemented in balanced mix design in Indiana.

Table 2.4 List of Performance Test Methods

Distress	Test Method	Testing Standard	Type of Test Result
Intermediate-Temperature Cracking	Direct Tension Cyclic Fatigue Test	AASHTO TP 107 AASHTO TP 133	Performance Index; Predicted Performance
	Disc-Shaped Compact Tension (DCT) Test	ASTM D7313	Performance Index;
	Flexural Bending Beam Fatigue Test	AASHTO T 321 ASTM D8273	Performance Index; Predicted Performance from empirical relationships
	Indirect Tensile Asphalt Cracking Test (IDEAL-CT)	ASTM D8225	Performance Index
	Illinois Flexibility Index Test (I-FIT)	AASHTO T 124	Performance Index
	Semi-Circular Bend Test (Louisiana method)	LADOTD TR 330 ASTM D8044	Performance Index
	Overlay Test	NJDOT B-10 Tex-248-F	Performance Index; Predicted Performance from empirical relationships
	N_{flex} Factor	AASHTO TP 141	Performance Index
Rutting	Asphalt Pavement Analyzer	AASHTO T 340	Performance Index
	Flow Number Test	AASHTO T 378	Performance Index; Predicted Performance from empirical relationships
	Hamburg Wheel-Tracking Test	AASHTO T 324	Performance Index
	Stress Sweep Rutting	AASHTO TP 134	Performance Index; Predicted Performance
	High Temperature Indirect Tension	None	Performance Index
	Rapid Shear Rutting Test (IDEAL-RT)	Draft ASTM Work Item (WK 71466)	Performance Index
Low-temperature Cracking	IDT Creep Compliance and Strength Test	AASHTO T 322	Performance Index; Predicted Performance
Mixture Toughness	Cantabro Test	AASHTO TP 108	Performance Index
Moisture Susceptibility	Tensile Strength Ratio	AASHTO T 283	Performance Index

2.4.2 Determination of performance failure criteria

The determination of the performance index threshold values is another important step in developing IPS or IPMD. In this article, methods to determine index limits in existing research are categorized into two primary approaches: using the performance of existing mixtures and using the existing performance index threshold from other testing methods.

2.4.2.1 Correlation with performance of existing mixtures

One practical way to determine the performance index criteria is to use the mixtures with known field performance. Meanwhile, comparing the testing results with field performance can validate the effectiveness of the test. In the study conducted by FHWA (Golalipour et al. 2021), the same set of asphalt mixtures were tested using multiple cracking testing methods, i.e., Cyclic Fatigue, I-FIT, IDAL-CT, Overlay, the Nflex Indirect Tension test, and the Cantabro abrasion Loss test. The testing results were compared with the observed cracking on the Accelerated Lane Facility (ALF). The asphalt mixture on each test lane varied in binder performance grades and contents of RAP and Reclaimed Asphalt Single (RAS). The comparison results showed that most of the cracking tests had good correlations with the amount of cracking observed on test lanes. Among the testing methods, the Sapp parameter from the Cyclic Fatigue test, CTIndex from the IDEAL-CT test, and the Nflex parameter from the Indirect Tension test showed the highest correlation (approximately 0.6).

Buttlar et al. (2020) suggesting using the field performance to determine performance criteria in a study to support the development of BMD specifications for the Missouri DOT. The study showed that DCT, I-FIT, and IDEAL-CT testing results had good correlations with the observed field performance, and the DCT test results yielded the highest R2. The results also indicated that the scores from the Pavement Condition Rating System (PASER) had higher correlations than the International Roughness Index (IRI). The PASER deterioration rate was then used to determine the thresholds for the performance tests. To calculate the threshold values for roads with different criticality levels, the relationship between the PASER deterioration rate was first obtained. The deterioration rate varied between 1.0 to 0, and lower rate indicated superior materials were used. The values of the performance indices corresponding to PASER deterioration rate at 0.4, 0.3, and

0.25 were used as the initial thresholds for mixtures targeting low, medium, and high criticality levels, respectively. The final values could be obtained after initial thresholds being adjusted with the consideration of aging and testing variability. A similar method was also applied in a study conducted for the development of BMD for Illinois (Buttlar et al. 2021).

In a study conducted in Louisiana, the threshold values for the HWTT rutting test and the SCB cracking test were determined based on the correlation between the testing results and field performance (Mohammad et al. 2016b). To set the rutting limit, an enclosed area under 6 mm of rutting in both field measurement and laboratory test were created for Level 2 traffic, as presented in Figure 2.4, and a 10 mm by 10 mm area were made for Level 1 traffic volume. The rutting limits of 6 mm and 10 mm were determined for Level 2 and Level 1 traffic, respectively.

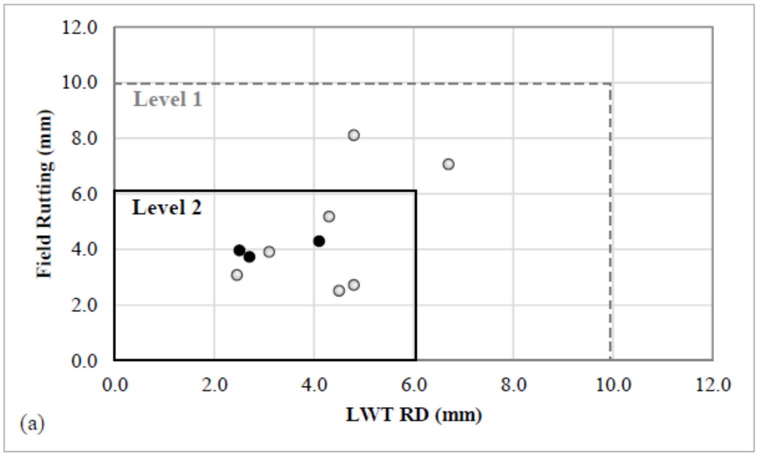


Figure 2.4 Tentative guidelines of laboratory rutting performance indicators (Mohammad et al. 2016b). (LWT RD: Loaded Wheel Tracking Rut Depth)

SHAs sometimes do not have sufficient field performance record for the tested mixtures when a new test method is introduced. In this case, Diffenderfer and Bowers (2019) proposed three approaches to determine performance criteria based on testing results from commonly used mixtures. Method 1 adopted the minimum value from the testing results as threshold, assuming all the mixtures would provide satisfactory performance since they had passed the existing volumetric criteria. This method would have the lowest risk for SHAs to initiate the BMD implementation. Method 2 used the average value of the testing results from all tested mixtures, which might lead to an immediate improvement by introducing BMD and yet might result in half of the existing mixtures to be rejected or re-designed. Method 3 considered the average value and the testing variability. The SHA can decide the suitable strategy to determine the criteria. For example, in

Virginia, since historically no severe rutting was reported, Method 1 for APA was adopted. However, because the cracking on pavements was the major concern, Method 2 was used to determine the IDEAL-CT index limit.

Lee et al. (2020) developed performance criteria for Indiana mixtures using methods similar to above-mentioned Method 2. The Indiana DOT intended to use the thresholds to exclude the poorest quality mixtures. The density function and cumulative distribution function of the reported values of the performance indices were plotted. The 10th and 20th percentile of the testing result of a series of mixtures were hereby calculated. The final performance threshold values considered the binder performance grade and the nominal maximum aggregate size (NMAS). For example, the FI value for 9.5 mm mixture containing PG 64-22 binder was 2.6, which was the 10th percentile limit among all the tested mixtures with the same NMAS and binder PG.

Etheridge et al. (2019) conducted a study to determine the threshold limits for the fatigue resistance index parameter, Sapp, obtained from the Cyclic Fatigue test for the Georgia DOT. Local mixtures with different binder grade, NMAS, and polymer modifiers were tested. The testing results were consistent with the observed performance of these mixtures, and the ranking of the mixtures with different design parameters were in line with the engineers' intuition. Good correlations were also observed between the Sapp values and predicted pavement fatigue cracking from FlexPAVE™ and the AASHTO Pavement ME Design Software. Based on the performance of the existing mixtures, the predicted pavement performance, and the SHA's requirements for mixtures at each traffic level, the threshold values of Sapp for the Georgia DOT were determined, as presented in Table 2.5. The values are different from the recommendations for the nation-wide applications (Wang et al. 2020).

Table 2.5 Recommended threshold values for S_{app} based on GDOT mixture selection criteria assuming 5% truck traffic and 1.17 ESAL factor (Etheridge et al. 2019)

Two-way ADT	Traffic level (million ESALs)	S_{app}	Mix type	Remarks
10,000–25,000	>4 and ≤10	>12	12.5-mm Superpave mix with PG 64-22 or PG 67-22 binder	For state routes and shoulders of interstate routes
25,000–50,000	>10 and ≤20	>15.5	12.5-mm Superpave mix with polymer-modified binder	For high-ADT state routes, interstate routes when recommended by GDOT, all flexible pavement interstate ramps, and all flexible pavement roundabouts

>50,000	>20	N/A	12.5-mm stone matrix asphalt	For interstate routes and for state routes when recommended by GDOT

2.4.2.2 Correlation with other testing methods

Another way to evaluate the effectiveness of the candidate performance tests is to compare the testing results from different tests when there is no sufficient field performance data. A high correlation between the testing results from two or more test methods can provide some confidence. One can determine the preliminary threshold index values for a new test based on the limits from an existing test.

There are multiple cracking tests available for performance construction specifications or mix design methods. Many research studies have compared the test methods and attempted to develop correlations among the testing results. Table 2.6 presents the correlations between testing results from different cracking test results reported in recent studies.

In some recent studies (Zhou et al. 2020 and Zhou et al. 2021), a correlation with R2 of 0.97 between the CTindex and the FI was presented. The limits for the CTindex obtained from the IDEAL-CT test were determined based on the existing threshold values of FI. As a reference, CTindex = 90 was believed to be equivalent to FI = 8.

As good correlations between the results from the IDEAL-CT test and the I-FIT test had been reported in many research studies, Al-Qadi et al. (2021) studied the mechanisms behind the two tests and presented a different opinion. The researchers pointed out that a good correlation between the CTIndex and the FI only exists in some conditions. With a notch induced in the SCB cracking test, the energy dissipates into the cracking propagation. In contrast, in the IDEAL-CT test, the energy is distributed between plastic deformation and cracking propagation. When plasticity is not dominating at low temperature and/or high loading rate, the good correlation between two tests can be observed. The study also used the Digital Image Correlation results to verify the mechanisms. The test results in the study showed that under different testing conditions, the R^2 between the CTindex and the FI varied between 0.56 and 0.04.

Table 2.6 Correlations between Cracking Tests in Recent Studies

Tests	R^2	Remark	Research
I-FIT and IDEAL-CT	0.61 and 0.81	• Ohio coarse asphalt concrete base mixture • Field cores from two projects • Over ten datapoints in each project	Garcia-Ruiz and Sargand 2021
IDEAL-CT and CPR* from Overlay	0.96	• Texas Superpave mixtures • limited number of tests and relatively high margin of error.	Al-Khayat and Epps Martin 2021
I-FIT and IDEAL-CT	0.74	• 12 field projects in Wisconsin	Abdalla et al. 2021
IDEAL-CT and S_{app} from Cyclic Fatigue	0.86	• 6 mixtures with different binder grades and contents of recycled materials • ALF test lanes	Golalipour et al. 2021
I-FIT and IDEAL-CT	0.89	• 6 mixtures with different binder grades and contents of recycled materials • FHWA ALF test lanes	Golalipour et al. 2021
I-FIT and IDEAL-CT	0.87	• 36 mixtures from Missouri	Buttlar et al. 2020
I-FIT and IDEAL-CT	0.97	• 9 datapoints from Texas mixtures	Zhou et al. 2020
IDEAL-CT and CPR from Overlay	0.90	• 18 datapoints from Texas mixtures	Zhou et al. 2020
I-FIT and IDEAL-CT	0.56	• 8 mixtures tested at low temperature and high loading rate	Al-Qadi et al. 2021
I-FIT and IDEAL-CT	0.04	• 5 mixtures tested at intermediate temperature	Al-Qadi et al. 2021

In terms of rutting test methods, new tests have also been introduced, for example, the IDEAL-RT test. In the study conducted by Al-Khayat and Epps Martin (2021), the testing results showed that the IDEAL-RT test and the HWTT could both effectively evaluate the rutting resistance of asphalt mixtures. A correlation with R2 of 0.92 between the measured APA rut depth and the shear strength from the IDEAL-RT test was reported in Zhou et al. (2020, 2021). To develop a BMD framework in Nebraska (Nsengiyumva et al. 2020), a new rutting test called Gyratory Stability test was proposed. A good correlation between the tested G-Stability from the Gyratory Stability test and the flow number was presented. The performance limit was then determined based on relationship between the two test results and the known flow number limits. The threshold values of prevalent performance tests in some current practices are presented in Table 2.7.

Table 2.7 Threshold Values of CT_{index} in Existing Research and Current Practices

SHA	Criterion	Testing Protocol	Air void (%)	Aging
Missouri (MoDOT 2021a,b)	CT_{index} between 32 and 60 for 100% pay, Superpave CTindex between 80 and 159 for 100% pay, SMA	ASTM D8225	7 ± 0.5	Short-term (AASHTO R 30)
Alabama (ALDOT 2020)	CT_{index} ≥55 (ESAL<1M, Mix design) CT_{index} ≥50 (ESAL<10M, Production Acceptance) CT_{index} ≥83 (ESAL<30M, Mix design) CT_{index} ≥75 (ESAL<1M, Production Acceptance) CT_{index} ≥110 (ESAL<10M, Mix design) CT_{index} ≥100 (ESAL<30M, Production Acceptance)	ALDOT 459	7 ± 1	Short-term (AASHTO R 30)
Oklahoma (Cross and Li 2019)	CTindex ≥ 80	ASTM D8225	7 ± 0.5	Short-term (AASHTO R 30)
Tennessee (Yin and West 2021)	CT_{index} ≥ 50 (State Routes, <10,000 ADT) CT_{index} ≥ 75 (State Routes, >10,000 ADT) CT_{index} ≥ 100 (Interstates and State Routes (controlled access)), <10,000 ADT)	ASTM D8225	7 ± 0.5	Short-term (AASHTO R 30)
Texas (TxDOT 2019)	Correlation between IDEAL-CT and Overlay test	Tex-250-F	7 ± 0.5	Short-term

2.4.3 Testing variability and uncertainty

Variability exists in all performance tests, and the uncertainty derived from the testing variability needs to be accounted in the performance specification/mix design. The testing variability of different performance tests has been reported in many studies. In the study conducted by Golalipour et al. (2021), mixtures from FHWA ALF lanes were tested using six performance testing methods: the Cantabro Mass Loss test, the N_{flex} test, the Overlay Test, the I-FIT test, the IDEAL-CT test, and the Cyclic Fatigue test. The coefficients of Variation (COV) are presented in Table 2.8**Error! Reference source not found.**. The test result showed that S_{app} parameter from the Cyclic Fatigue test and the critical fracture energy parameter from the Overlay Test had the lowest COV. The FI parameter from the I-FIT showed the highest COV among all the tests. Garcia-Ruiz et al. (2021) compared the testing variability of the I-FIT test and the IDEAL-CT test. The CT Index from the IDEAL-CT test showed 30% less variability than the FI from the I-FIT test in the study.

To understand and estimate the testing variabilities in different laboratories, before the full deployment of the BMD, the Virginia DOT started round robin to evaluate the testing repeatability. Forty-one laboratories from the agency, contractor, and independent testing labs participated. Two different mix designs were used, and 46 sets of test specimens for each design were distributed. The Phase 1 study aimed to evaluate the IDEAL-CT test results from different labs, followed by the Phase 2 study starting in 2021 (Diefenderfer et al. 2020, VAA 2021).

Table 2.8 Coefficient of Variation (COV) of Performance Tests (Golalipour et al. 2021)

Mixture	Cantabro Mass Loss	N_{flex}	OT: Crack Resistance Index	OT: Critical Fracture Energy	FI	CT_{Index}	S_{app}
Lane 3	20%	18%	17%	4%	28%	30%	3%
Lane 5	5%	20%	19%	2%	29%	7%	3%
Lane 6	18%	22%	6%	5%	30%	14%	3%
Lane 7	18%	31%	11%	1%	29%	40%	6%
Lane 8	1%	31%	7%	4%	34%	1%	1%
Average	12%	24%	12%	3%	30%	18%	3%

Another way to estimate the testing uncertainty was demonstrated by Ding et al. (2020). The researchers computed the uncertainty in the cyclic fatigue testing results given the testing variabilities in the precedent dynamic modulus measurements and the fatigue test. The Bayesian Inference-based Markov Chain Monte-Carlo method was adopted. In the analysis, the uncertainties from the previous step (i.e., dynamic modulus test) propagated into the following step (i.e., the fatigue test, in this case), and the interactions from different steps were accounted in the final calculation. The 95% credible interval and 95% prediction interval were calculated when different number of testing replicates were used in the analysis. Four replicates of the Cyclic Fatigue test were recommended considering the balance between the acceptable uncertainty of fatigue prediction and the testing efficiency.

A completed performance specification/mix design method should consider the testing variabilities and uncertainties. In the PRS developed by Caltrans (Harvey et al. 2014), the specification limits were determined based on the 95% confidence interval for the given measured material properties. Similarly, Buttlar et al. (2020) performed cracking tests (i.e., DCT test, I-FIT test, and the IDEA-CT test) to develop a local BMD method. It was reported that while the average DCT COV was 19.5% for field section evaluations, the I-FIT test had a much higher average COV

(52.2%). The recommended threshold values for the performance tests included two times of standard deviation to accommodate the testing variabilities.

2.4.4 Aging conditioning and sampling for design and QA

Aging is one important factor for mix performance evaluation during design and QA. The short-term and long-term aging conditioning should be performed to simulate the aging effects during construction and service. However, the conditioning is time consuming. The short-term aging may take 2 to 4 hours while the long-term aging usually requires at least eight hours (Zhou et al. 2021). Based on the binder chemistry and rheology study in the NCHRP 9-54 project, depending on the climate conditions, one may need to oven condition loose mixes for 24 to 696 hours at 95 °C (Kim et al. 2018) to simulate the long-term field aging. Moreover, during production, the aging level of the mix samples may vary depending on the sampling time, storage time in the silo, hauling time to the construction site, and so forth. Thus, it is critical to develop methodologies to incorporate aging factors to implement performance mix design and construction specification.

Researchers have provided several potential strategies to consider the aging effects. For example, Zhou et al. (2020 and 2021) proposed a short-term aging protocol with loose mixture conditioned at compaction temperature for 2 hours for both cracking and rutting tests during mix design and QA and a mid-term aging procedure (20 hours at 100°C) following the short-term aging treatment before the loose mixture were reheated and compacted to fabricate cracking testing samples for performance verification during mixture acceptance. To minimize the inconsistencies of aging levels in mix samples, the Missouri DOT (MoDOT) requires the QC and QA samples for IDEAL-CT cracking tests to be compacted in the plant laboratory without reheating the loose mixes in their recent specification (Balanced Mix Design Performance Testing for Job Mix Approval NJSP-21-08A).

In addition, since the performance tests usually have longer turnaround time than the conventional volumetric tests, the sampling method and the testing frequency should be well-planned. Williams et al. (2004) conducted a thorough research for the development of a performance specification, and a study about sampling methods was included. Different sampling methods, including sampling in a haul truck, sampling in a haul truck from a platform, and sampling on roadway behind a paver, were compared, and the sampling procedure for each method was documented. It was found that the mixture test results correlated better with tank-sampled binder than with the truck recovered samples. The testing results also indicated that binder aging was still occurring during the transport and laydown of the mixture.

Tremendous research effort has been spent to understand the difference between asphalt samples prepared by different methods (lab produced vs. plant produced). Mohammad et al. (2016a) found that the process-based factors (i.e., return of baghouse fine, delay in specimen fabrication, aggregate absorption, aggregate hardness, and stockpile moisture content) did not have significant effects on the volumetric properties and mechanical properties. However, the sample preparation methods (i.e., plant-mix field-compact (PF), lab-mix lab-compact (LL), and plant-mix lab-compact (PL)) had impact on the mechanical properties of the mixtures. The PF samples were significantly softer than the LL and PL samples at the same air void level, and PF samples consistently yielded higher rut depth in the loaded-wheel test. The study proposed conversion factors to convert the tested rut depth with one sample preparation method to rut depth with other preparation methods. The finding was confirmed by Liu et al. (2017). Liu et al. (2017) also concluded that the variability in volumetric property measurements were found generally lower than in the mechanical property measurements. Daniel et al. (2018) found that the reheating process during the preparation of PL samples had a great impact on the mixture properties. Similar trend was also reported by Al-Khayat and Epps Martin (2021). In addition, several research studies have indicated that the measured fatigue cracking resistance of LL mixtures was consistently higher than that of plant-mixed mixtures (Johanneck et al. 2015, Newcomb 2018, Lee et al. 2019). The sampling position and testing frequency in existing research and current practices are presented in Table 2.9**Error! Reference source not found.**.

Table 2.9 Sampling Position and Testing Frequency in Performance Specifications

State	Sampling Position	Testing Frequency
Minnesota (Newcomb 2018)	Plant	every 2000 tons (wearing course)
Missouri (MoDOT 2021a)	Plant	every 10,000 tons
Alabama (ALDOT 2020)	Plant	once per day (agency) every 700 tons (contractor)
Virginia (Diefenderfer and Bowers 2019)	Plant	every 1000 tons (agency) every 500 tons (contractor)
California (Yin and West 2021)	Plant	every 10,000 tons or once per project (HWTT) three specimens per day (flow number and I-FIT)
Illinois (IDOT 2021)	Behind paver	once per day or every 1000 tons
Louisiana (LDOTD 2016)	Plant	every 2000 tons
Louisiana (Mohammod et al. 2016b)	Field core	25 cores per lot
Vermont (VTrans 2019)	Plant	every 3000 tons

2.5 Summary and recommendations

In this state-of-the-art review, the efforts in ensuring the durability and performance of asphalt mixture in mix design and production are documented. The development of the performance specification and performance mix design is summarized. The challenges and solutions in incorporating performance tests in asphalt productions are discussed. The chapter aims to provide hints to the asphalt community in developing holistic methodologies and integrating quality assurance in performance mix design. The summary and future recommendations are stated as follows.

2.5.1 Under the framework of volumetric design

- Some states have proposed to increase the VMA by 1.5 to 2.0% and remove the upper limit of VFA. Some agencies increased the design air void or increased the N_{des} aiming to gain more compaction effort during field placement. Some others decreased the design air void or decreased the N_{des} to create more space for binder.
- SHAs should be careful when adjusting the volumetric requirements. The interactions

between volumetric variables should be thoroughly considered. Increasing the VMA can theoretically generate more room for binder; however, the adjustment may lead to changes in gradation and mixture compactability.

- It is difficult to use volumetric-based design method only to accommodate the usage of recycled materials and support contractors to apply new technologies in asphalt mixtures.

2.5.2 Development of performance specification and performance mix design

- The definition and history of different types of construction specifications are summarized in this article. The relationships among all types of construction specifications are illustrated in Figure 2.1**Error! Reference source not found.**.
- The chapter categorizes performance specification into index-based performance specification and predictive performance specification which include performance-related specification (PRS) and performance-based specification (PBS).
- Predictive performance specifications utilize fundamental material engineering properties to predict pavement performance and use the difference between the as-design and as-built pavements to determine the pay adjustments. While no true PBS exists in practice due to difficulties in measuring the fundamental material properties. PRS is one solution to the concern, which uses AQCs to estimate fundamental material properties and predict pavement performance. Challenges to deploy PRSs include to effectively quantify and encompass the risk and reliability, to provide training and instructions for SHAs and contractors, to develop local material database and local model calibration coefficients, and to simplify the performance tests.
- Index-based performance specifications (IPSs) are QA specifications with simple performance tests incorporated. Compared to predictive performance specifications, IPSs are more intuitive and can adopt performance tests that can be completed in a timely manner. However, selecting the suitable testing methods and determining appropriate index threshold limits are major challenges in IPSs. Besides, IPSs have limited capability to consider project-specific information such as traffic, structure, and climate conditions.
- Four approaches in balanced mix design/index-based performance mix design and one predictive performance mix design method are demonstrated. The five approaches are

compared based on six evaluation criteria proposed in this chapter.

2.5.3 Challenges and solutions in developing and deploying performance specification / mix design

- Selecting the testing method is one important step to develop a BMD. SHAs should consider key factors such as users' experience, availability of existing data, availability of testing equipment, cost of the test, turnaround time, and the effectiveness of the test when selecting a performance test.
- One major challenge to develop an index-based performance specification is to determine the appropriate index threshold limit. This article provided two approaches: using the performance of existing mixtures and using the using the existing performance index threshold from other testing methods.
- When there is no sufficient field performance data, preliminary threshold values can be determined using testing results from existing mixtures. Three approaches to select the limits are listed in the chapter.
- Many researchers have reported good correlations between different performance tests. The testing results from the I-FIT test and the IDEAL-CT test were found to have high correlations in several studies. However, some researchers (Al-Qadi et al. 2021) pointed out that the good correlation could only exist under limited conditions. The relationships between two tests can be used to determine the threshold values for a new test method.
- Different performance tests have different levels of COV. The testing variability should be accounted in the determination of the threshold limits.
- Researchers have found the lab-mix lab-compacted samples generally performed better than the plant-mix samples. Reheating the plant-mix loose mixes during sample preparation can have great impact on the testing results. Sampling position and testing frequency in current practices are presented in this chapter.

2.5.4 Future Research Directions

- Predictive performance specification/mix design is more complicated but may have long-term benefits compared to the index-based methods. Researchers may continue to simplify the testing methods and the model calibration process to further practicalize the framework.

- The risk and reliabilities in using the index-based performance/mix design should be quantitatively evaluated. The pay adjustments determined based on the index values should be further justified with systematic research studies.
- Comprehensive material testing database and pavement performance database should be established. The performance of the mixtures with performance testing results should continue to be monitored. The long-term benefits of using performance specification/mix design can be evaluated.
- To further simplify the performance specification/mix design, one way is to accurately predict the performance testing results from simple measurable variables. The new technologies like the machine learning algorithm can contribute to this direction.

Chapter 3 Development of a coherent framework for BMD and QC-QA framework

3.1 Introduction

Asphalt mixtures are becoming increasingly complex. In the last 10 years, the use of reclaimed asphalt pavements (RAP), recycled asphalt shingles (RAS), fibers, and rejuvenators in some cases, has become the new norm. Furthermore, asphalt binder sources, refineries, and modification techniques (polyphosphoric acid, re-refined engine oil bottom, recycled plastics, and others) are dynamically changing the landscape for mixtures. Given ever-changing components of asphalt mixtures, many state departments of transportation (DOTs) are in the process of developing or preliminarily implementing some type of performance specification for asphalt mixtures to ensure mixture durability. For example, many DOTs initiated the adoption of balanced mix design (BMD) approaches (Zhou et al. 2007, Bennert 2011, Zhou et al. 2014, Mohammad and Cooper 2016, Buttlar et al. 2016, Ozer and Al-Qadi 2018, West et al. 2018, Newcomb and Zhou 2018). BMD is a crucial step forward in designing a well-performing mix with balanced rutting and cracking resistance, but it is not the whole performance specification. Another critical component of a performance specification is quality control and quality acceptance (QC/QA) testing during the production process. Regardless of how well a mixture is designed in a laboratory, if the mixture quality is not properly controlled during production, the mixture performance in the field could be jeopardized. Current production QC/QA testing focuses on three major characteristics of asphalt mixtures: asphalt content, aggregate gradation, and laboratory-compacted density. These characteristics are important, but sometimes they are not directly related to mixture performance. For example, during the production process, one may have to replace one source of PG64-22 asphalt binder with another binder source due to supply shortage. This replacement of the binder source often has no influence on asphalt binder content, aggregate gradation, and volumetric properties, so the produced mixture will pass all three QC/QA tests. However, such replacement could have a significantly negative influence on cracking resistance (Mogawer et al. 2019). To ensure what is produced at the plant is similar to what was originally designed in the laboratory, the same (or similar) performance tests used for BMD are preferred for production QC/QA testing. However, some performance tests are suitable for BMD, but they may not be practical for QC testing. For example, the Hamburg Wheel Tracking Test (HWTT) has been widely used by many

DOTs in mixture design to ensure adequate rutting/moisture damage resistance. However, the long testing period of HWTT prevents it from being an efficient test to implement for production QC. Therefore, the main objective of Task 2 was to recommend a coherent framework for both laboratory BMD and production QC/QA, which includes the newly developed concept of mid-term aging.

This chapter first presents the envisioned framework, which includes four major components: (1) volumetric mix design, (2) performance evaluation of multiple asphalt contents and selection of a balanced binder content, (3) performance verification of the selected balanced asphalt content, and (4) a QC/QA testing plan and associated acceptance criteria. The following three areas are then further discussed:

- Selection of performance tests for the framework.
- Laboratory aging (or conditioning) protocols for preparing specimens.
- Performance tests' criteria and strategies for meeting those criteria.

In addition, a case study is presented to demonstrate the whole process of the coherent BMD/QC/QA framework.

3.2 Envisioned coherent BMD/QC/QA Framework

Designing stable and durable asphalt mixtures has been pursued by different methods for decades. Generally, mixture stability (or rutting resistance) is controlled through a strength test, such as Marshall stability, while mix durability (e.g., cracking resistance) is often ensured by adequate asphalt binder content through volumetric requirements—for example, air voids and voids in mineral aggregate. The BMD concept was not introduced until 2007 when Zhou et al. (2007) employed two performance tests—the Hamburg wheel tracking test (HWTT) and Texas Overlay Test (OT)—to evaluate rutting and cracking resistance of asphalt mixtures under multiple asphalt contents selected based on volumetric mix design. Three specific features of the original BMD concept are as follows (Zhou et al. 2007): (1) allowing air void variation from 2 to 5 percent when selecting asphalt contents for performance evaluation, (2) multiple (3 or 4) asphalt contents selected for performance evaluation, and (3) relying on rutting/moisture damage and cracking performance tests and associated criteria to define a balanced asphalt content zone/range meeting both rutting and cracking requirements. Since then, different forms of BMD have been explored. In 2015, federal highway administration (FHWA) formed a BMD task force, and the task force identified three potential approaches to the use of BMD: (1) volumetric design with performance testing validation, (2) performance-modified volumetric mix design, and (3) performance design. Each approach has its advantages and limitations. Comparatively, Approach 1 is the easiest one to understand and implement, but it is limited to the evaluation of a single selected asphalt content with 4 percent air voids. A necessary enhancement to Approach 1 is to test three asphalt contents with the same aggregate blend but that correspond to air voids varying from 2 to 5 percent, as proposed by Zhou et al. (2007). Approach 2 is a performance-modified volumetric mixture design. This approach is similar to the first approach in that it starts by determining the optimum binder content using the Superpave volumetric design method but subsequently focuses on meeting mixture performance test criteria. The mixture design binder content and/or proportions can be adjusted to accommodate the performance test requirements. The final design may not be required to meet all the volumetric Superpave criteria. The accuracy of the performance tests and performance models in Approach 2 becomes extremely important since final mix design is dictated by performance tests and models rather than volumetric properties. Approach 3 is a reverse version of Approach 1. Performance requirements are met first, and then the mixture volumetric properties are determined. Implementing Approach 3 requires mindset changes that are not intuitive to most asphalt mix technologists. These design approaches have been attempted with varying degrees of success in the past. However, only limited work has been done for production QC/QA testing within the BMD framework. Building on existing knowledge and experiences with BMD and QC/QA testing at asphalt plants, the research team developed a coherent framework for BMD/QC/QA (see **Error! Reference source not found.**), including four major components: (1) volumetric mix design, (2) performance evaluation at multiple asphalt contents to select the balanced asphalt content, (3) performance verification at the balanced asphalt content, and (4) development of QC/QA testing plan and acceptance criteria. Each of the components is briefly described below.

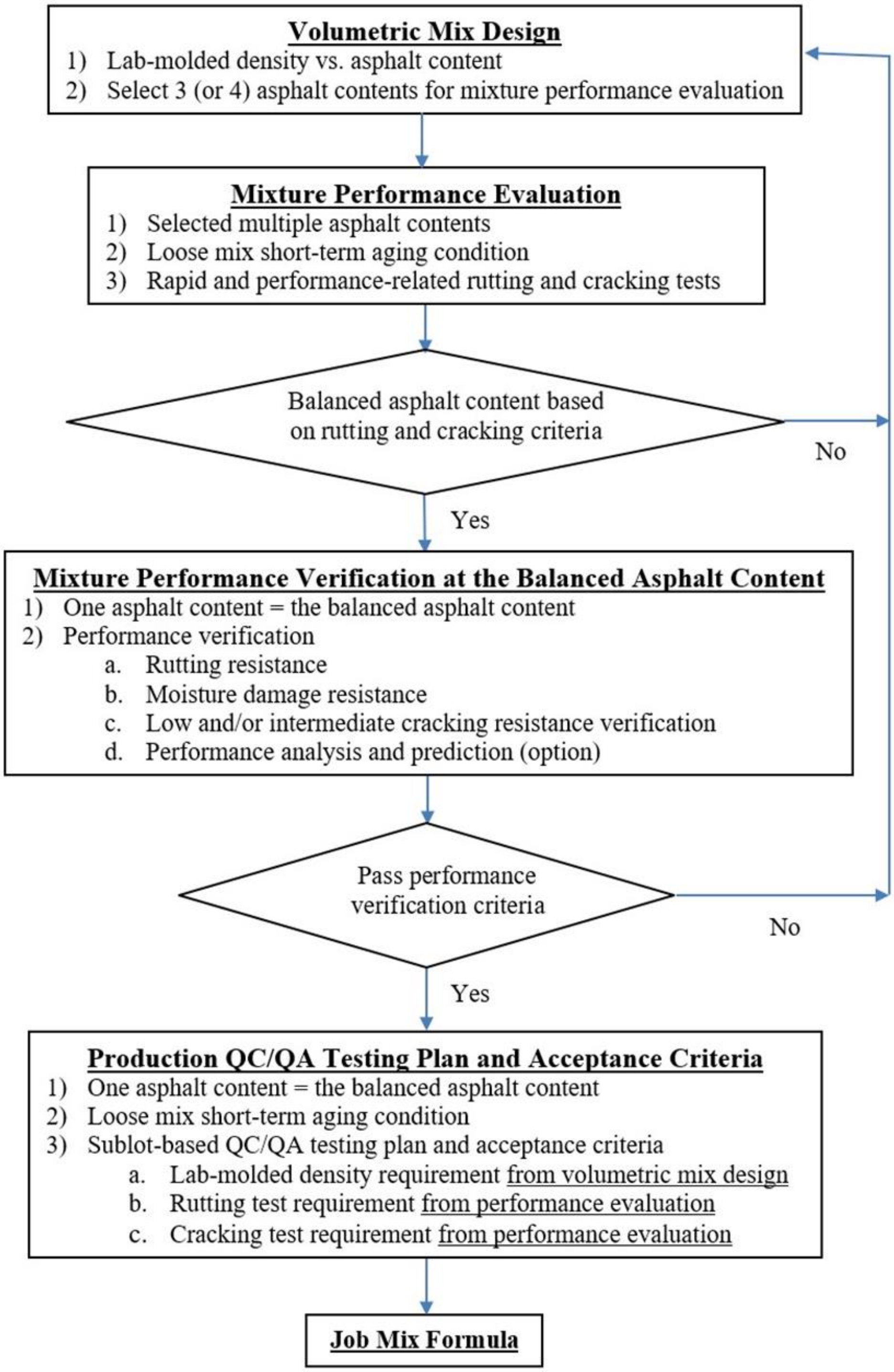


Figure 3.1 Envisioned Coherent Framework for BMD/QC/QA

3.2.1 Volumetric mix design

Volumetric properties of asphalt mixtures have been the backbone of asphalt mix design methods and production QC/QA testing for decades. Not only are current mix designers very familiar with the volumetric designs, but they have also accumulated much useful experience that can help implement a coherent BMD/QC/QA framework. In the envisioned framework, the volumetric mix design will serve the following two purposes:

- The next-step selection of multiple asphalt binder contents for performance evaluation: Based on longstanding practice, the asphalt content at 96 percent lab-molded density (or 4 percent air voids) is often selected for performance evaluation (as in Approach 1). However, such a practice may inhibit mix designers from innovation and developing a good performing mixture. Moreover, the selection of 96 percent lab-molded density is based on the assumption that the density of the asphalt mixture ultimately becomes 96 percent at the end of its life—from the initial 92–93 percent construction density—after years of traffic densification. This process is also known as terminal density (96 percent). However, NCHRP 9-09(1) reported that most Superpave mixes never reach 96 percent density (Prowell and Brown 2007). Historically, design densities can vary from 95 to 97 percent in the Marshall mix design method. Furthermore, various field test sections designed with a laboratory-molded density ranging from 96.5 to 98 percent were constructed in Texas under Projects 0-6092, 0-6614, and 0-6738, and no rutting was observed (Zhou et al. 2011, Zhou et al. 2013, Im et al. 2015). Thus, the envisioned framework is not limited to 96 percent lab-molded density for selecting asphalt binder content. Instead, at least three asphalt contents will be selected at lab-molded densities ranging from 95–98 percent for performance evaluation.
- Defining lab-molded density (or air voids) requirement for production QC/QA: Lab-molded density is still a critical component of performance-related QC/QA testing (or specification). Although the final asphalt content is not determined at the volumetric mix design stage of this framework, the density versus asphalt contents curve obtained here will be used to define the acceptance lab-molded density for production QC/QA after the final design asphalt content is verified later.

A later section of this chapter demonstrates these two purposes with actual mix design data, including the development of the QC/QA density requirement.

3.2.2 Mixture performance evaluation

The performance evaluation component of this framework evaluates rutting and cracking resistances of asphalt mixtures at multiple (3 or 4) asphalt contents rather than one asphalt content at 96 percent lab-molded density so that a balanced asphalt content is selected (or optimized) within a balanced zone. Since multiple asphalt contents are involved, many specimens are required. Thus, a set of simple, rapid, and performance-related rutting and cracking tests are preferred. Further, the same set of rutting and cracking tests will be employed during the production QC/QA testing. In this manner, the results of mixture performance evaluation in the lab mix design stage can be used to develop plant production QC/QA acceptance criteria. Moreover, since time is limited for QC/QA testing, short-term aging is desired at this stage versus long-term aging.

3.2.3 Mixture performance verification at the Balance Asphalt Content

Different from the mixture performance evaluation described above, the performance verification focuses on verifying mixture properties at the selected balanced asphalt content rather than multiple asphalt content, specifically verifying those mixture properties not evaluated in the stage of the mixture performance evaluation, such as moisture susceptibility and cracking resistance at intermediate temperature or low temperature after long-term aging. Moreover, since only one asphalt content is considered here, some repeated loading tests with relatively long testing time (such as OT, flexural beam fatigue test, Asphalt Mixture Performance Tester cyclic fatigue test, etc.) can be employed. If preferred, field performance of the asphalt mixture at the balanced asphalt content can be predicted using software such as the American Association of State Highway and Transportation Officials (AASHTO) AASHTOWARE Pavement ME, CalME (Ullidtz et al. 2010), TxME (Hu et al. 2014), or FHWA Flexpave™ (Wang et al. 2018).

3.2.4 Production QC/QA testing plan and acceptance criteria

The production QC/QA testing plan and acceptance criteria are a critical component of the framework. To overcome the deficiencies of the current QC/QA testing plan focusing on lab-molded density, asphalt content, and aggregate gradation, Zhou et al. (2020) developed a performance-related QC testing plan relying on three mixture properties: lab-molded density, rutting resistance, and cracking resistance. The acceptance criteria of these three mixture properties can be determined directly from previous volumetric mix design and performance evaluation stages, respectively. In this fashion, not only are the BMD and the production QC/QA interconnected, but the same tests measuring lab-molded density, rutting and cracking resistance, and associated criteria employed for selecting the balanced asphalt content can be used for the production QC/QA testing, which will save time and resources. It is advantageous to use the same asphalt mixture-aging condition(s) for the production QC/QA testing as was used for the volumetric mix design and QC/QA performance evaluation.

In short, a coherent BMD/QC/QA framework was established. However, the research team believed some details were still missing: (a) laboratory tests for the performance evaluation and verification stages, (b) laboratory aging (or conditioning) protocols for preparing specimens for volumetric mix design, mixture performance evaluation and verification, and production QC/QA testing, and (c) performance test criteria and strategies to meet those criteria.

3.3 Selection of laboratory performance tests

Currently, all forms of cracking—reflective, top-down, low temperature, and fatigue—are a major concern for state highway agencies. Meanwhile, to perform well in the field, an asphalt mixture must have good resistance to rutting and moisture damage, which are traditional mix design criteria. Thus, BMD, at minimum, should have three performance tests: (1) cracking, (2) rutting, and (3) moisture damage. As shown in **Error! Reference source not found.**.1, this framework requires laboratory tests for the performance evaluation and verification and production QC/QA testing. Due to different requirements for each application (performance evaluation, verification, and QC/QA), two sets of performance tests were selected, as described below:

3.3.1 Laboratory Tests for the Mixture Performance Evaluation and the Production QC/QA

As discussed previously, two types of performance tests—rutting and cracking tests—are needed for both the BMD performance evaluation and the production QC/QA testing. Many different laboratory tests, as shown in **Error! Reference source not found.**3.1, have been developed for characterizing cracking and rutting properties of asphalt mixtures in the literature (Zhou et al. 2016, West et al. 2018, Hajj et al. 2019, Zhou et al. 2020).

Table 3.1 Common Performance Tests for Cracking, Rutting, and Moisture Susceptibility

Cracking test		**Rutting test**		**Moisture damage test**	
ASTM D7313 Disk-shaped compact tension (DCT) test		ASTM D6927 Marshall stability test		AASHTO T283 Tensile strength ratio (TSR) test	
AASHTO TP105 Semi-circular bend (SCB)-low temperature		AASHTO T324 HWTT		AASHTO T324 HWTT	
ASTM D8044 SCB-critical strain energy release rate (Jc) test		AASHTO T340 Asphalt pavement analyzer (APA) test			
AASHTO TP124 [24] Illinois flexibility index test		AASHTO TP79 Flow number test			
IDT-University of Florida method or AASHTO T322		AASHTO T320 Superpave simple shear test			
Tex-248-F OT		AASHTO TP116 iRLPD test			
AASHTO T321 Beam fatigue test		AASHTO TP 134 Stress sweep rutting test			
ASTM D8225 IDEAL-CT		High-temperature IDT strength test (Christensen et al. 2002)			
AASHTO TP107 AMPT cyclic fatigue test		ASTM WK71466 IDEAL-RT shear strength test (Zhou et al. 2019)			

Each test has its own features and applications, and not all performance tests are suitable for production QC testing. When selecting tests for both the BMD performance evaluation and production QC/QA testing, the research team considered many aspects (such as sensitivity, repeatability, etc.) and gave special attention to the factors listed below:

- Use of same set of laboratory tests for both the BMD performance evaluation and the production QC/QA testing: Zhou et al. (2020) established a hybrid approach wherein different sets of laboratory tests are used for BMD performance evaluation and the production QC/QA testing, respectively. In order to establish QC/QA acceptance criteria, extra work was needed to develop the correlations between the BMD performance tests and the production QC/QA tests. To avoid this extra effort and reduce potential errors induced by correlations between different laboratory tests, the research team recommends the use of the same set of laboratory tests for both the performance evaluation and the production QC/QA testing so that the test data collected during performance evaluation can be used for establishing the production QC/QA acceptance criteria.

- Test correlates well with field performance: The selected laboratory performance tests must correlate with observed field performance. Otherwise, the tests should not be chosen, regardless of how rapid, simple, or repeatable they may be.
- Simplicity and rapidity: The production QC testing is performed at asphalt plants that are often located in remote areas where sophisticated laboratory testing equipment or sample preparation tools (such as a saw or drill/core machine) are often not available. Thus, those performance tests not requiring instrumentation, cutting/notching, gluing, or coring/drilling are favored. Quick turnaround is another preferred feature for production QC testing. These preferences often exclude many research-level performance tests from consideration, although these tests could be used for BMD performance verification.

Considering the three major factors, the research team selected one cracking and two rutting performance tests for BMD performance evaluation and production QC/QA testing, as noted below.

- One cracking test:

IDEAL-CT: IDEAL-CT is standardized in ASTM D8225. It is run at the loading rate of 50 mm/min at 25°C. This test is often completed within 2 min after taking a specimen out of a conditioning chamber (e.g., water bath). The IDEAL-CT uses the Cracking Tolerance Index (CTIndex) as its cracking parameter. The larger the CTIndex, the better the cracking resistance of the mixture. The IDEAL-CT has a good correlation with field cracking performance (Zhou et al. 2017, West 2019). A minimum of three specimens (preferably more) 150 mm in diameter and 62 mm in height are molded at 7±0.5 percent air voids using a Superpave gyratory compactor. The IDEAL-CT is a rapid, simple, repeatable test that is sensitive to asphalt mix composition (aggregate, binder, recycled materials) and aging conditions. It is recommended that the IDEAL-CT be used for the BMD cracking performance evaluation and the production QC/QA testing to ensure the mix's adequate cracking resistance. The IDEAL-CT is being adopted by 14 DOTs as their cracking test (West 2020)

- Two rutting tests:
  - IDEAL-RT: IDEAL-RT is currently being balloted in ASTM WK71466: Standard Test Method for Determination of Rutting Tolerance Index of Asphalt Mixture Using the Rapid Rutting Test. It is run at a loading rate of 50 mm/min at the high temperature of 50°C and is often completed within 2 min after taking a specimen out of the conditioning chamber (e.g., water bath). The IDEAL-RT uses the Rutting Tolerance Index (RTIndex) as its rutting parameter. The larger the RTIndex, the better the rutting resistance of the mix. The IDEAL-RT has a good correlation with field rutting performance (Zhou et al. 2019). Three specimens with 150 mm in diameter and 62 mm in height are molded at 7±0.5 percent air voids using a Superpave gyratory compactor. The IDEAL-RT is a rapid, simple, repeatable test that is sensitive to asphalt mixture composition (aggregate, binder, recycled materials) and aging conditions. It is recommended that the IDEAL-RT be used for the BMD rutting performance evaluation and then for production QC/QA testing to ensure the mixture's adequate rutting resistance.
  - High-Temperature Indirect Tensile (IDT) Strength Test: Contrary to the IDEAL-RT directly measuring shear strength, the high-temperature IDT strength test measures the cohesion component of the shear strength of asphalt mixtures (Christensen et al. 2002), although it cannot capture the friction angle that also contributes to the shear strength of asphalt mixes. The NCHRP Project 9-33 recommended this test and some preliminary acceptance criteria for evaluating rutting resistance during the mix design (Advanced 2011). Recently, Bennert (2011) and Bennert et al. (2020) employed this test for the production QC tool, and it was conducted at 50 mm/min at a test temperature 10°C lower than the 50 percent reliability, 7-day average maximum pavement temperature obtained from LTPPBind Version 3.1. Overall, the high-temperature IDT test is rapid, simple, and repeatable, and has acceptable correlation with field rutting performance (Advanced 2011). Thus, this test method is also selected as an alternative rutting test for the BMD rutting performance evaluation and the production QC/QA testing.

3.3.2 Laboratory Tests for the Mixture Performance Verification

As described previously, the mixture performance verification needs to address (a) moisture susceptibility, (b) intermediate temperature cracking resistance after long-term aging, (c) low-temperature cracking after long-term aging, and (d) rutting resistance under repeated loading test(s). Furthermore, recognizing that each DOT may have their own preferred test(s) for each specific distress, the research team provided multiple options for the mixture performance verification testing, as detailed below:

- Moisture susceptibility test: As listed in Table 3.1, the two most often used moisture susceptibility tests are AASHTO T324: HWTT and T283: Lottman (or TSR) test. Both tests are recommended for verifying performance of the asphalt mixture at the balanced asphalt content, and users can choose either one based on their experiences and preference. The research team selected the HWTT for verifying moisture susceptibility in the demonstration case study because the HWTT can also serve as the rutting verification test.
- Rutting verification test: Although many rutting tests are available, most DOTs are using either AASHTO T324: HWTT or T340: APA, as noted by West et al. (2018). Thus, both tests are recommended for verifying the performance of asphalt mixture at the balanced asphalt content, and users can choose either one based on their experiences and preference. The research team selected the HWTT for verifying rutting resistance in the following case study since the HWTT can serve as the moisture susceptibility test as well.
- Cracking verification test at intermediate temperature: Various suitable cracking tests exist, as listed below:
    - AASHTO TP124: SCB-IFIT used by Illinois.
    - ASTM D8044: SCB-Jc used by Louisiana.
    - AASHTO T321: Flexural beam fatigue test used by California.
    - Tex-248-F: OT used by Texas and New Jersey.
    - ASTM D8225: IDEAL-CT used by Virginia, Oklahoma, Kentucky, Maine, and more.
    - AASHTO TP107: AMPT cyclic fatigue test recommended by FHWA.

Each DOT can choose their own cracking test to ensure adequate cracking resistance of asphalt mixtures after long-term aging. The research team chose IDEAL-CT to verify cracking resistance after mid-term aging, defined later in the following case study, due to its simplicity and good correlation with field performance.

- Low-temperature cracking verification test: Low-temperature cracking can be a serious concern for northern states or cold climate areas. Laboratory tests developed to characterize low-temperature cracking resistance of asphalt mixes include AASHTO TP105: SCB-low temperature; AASHTO T322: low-temperature IDT; ASTM D7313: DCT; and more. However, most of these methods are generally used as research-level performance tests. ASTM D7313: DCT was used by Minnesota DOT in a few pilot projects (Dave et al. 2019). Thus, the research team recommends ASTM D7313 DCT for verifying the low-temperature cracking performance of asphalt mixtures at the balanced asphalt content. The resulting fracture energy can also be used in conjunction with a software (Illi-TC) to further enhance pavement low-temperature cracking predictions.

In summary, for rutting, cracking, and moisture susceptibility, a series of laboratory tests with a high flexibility of fitting different needs and preferences of DOTs are recommended for the framework. This chapter specifically focuses on three laboratory tests (IDEAL-CT, IDEAL-RT, and HWTT) as the primary performance tests for the framework (Figure 3.2), and those tests will also be used in the case study:

- IDEAL-CT and IDEAL-RT for the BMD mixture performance evaluation and the production QC/QA testing.
- HWTT for BMD rutting and moisture susceptibility performance verification.
- IDEAL-CT for BMD cracking performance verification after mid-term aging.

Figure 3.3 shows another index-based framework with different performance tests. In contrast, Figure 3.4 illustrates a performance-based framework where pavement performance is simulated during mix design stage. Note that each approach has its advantages and disadvantages. DOTs should select different approaches in accordance with their own specific conditions and applications.

After selecting performance tests, the identification of appropriate asphalt mixture conditioning protocols for loose mixtures to mimic field aging is an important next step because aging protocols have a significant impact on mixture performance, and consequently, mixture acceptance criteria depend on how the mixture is aged. Thus, the next section discusses aging protocols.

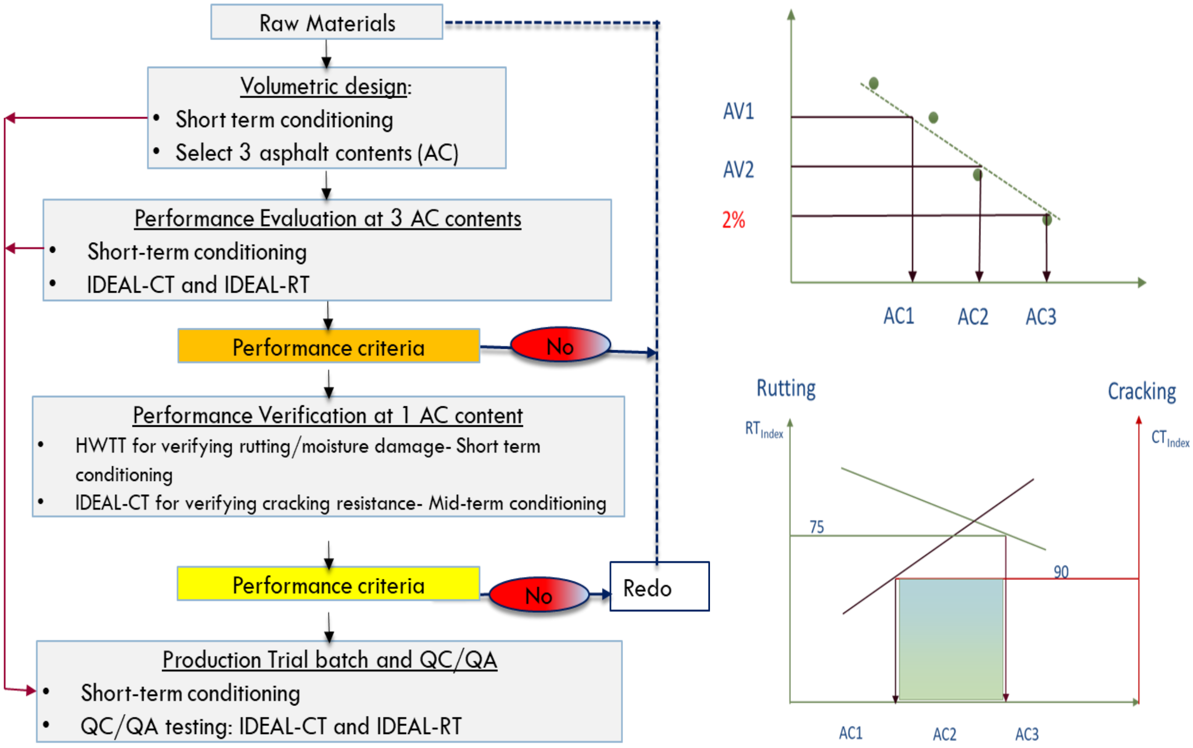


Figure 3.2 Coherent Framework for BMD/QC/QA: Example One with Index Tests

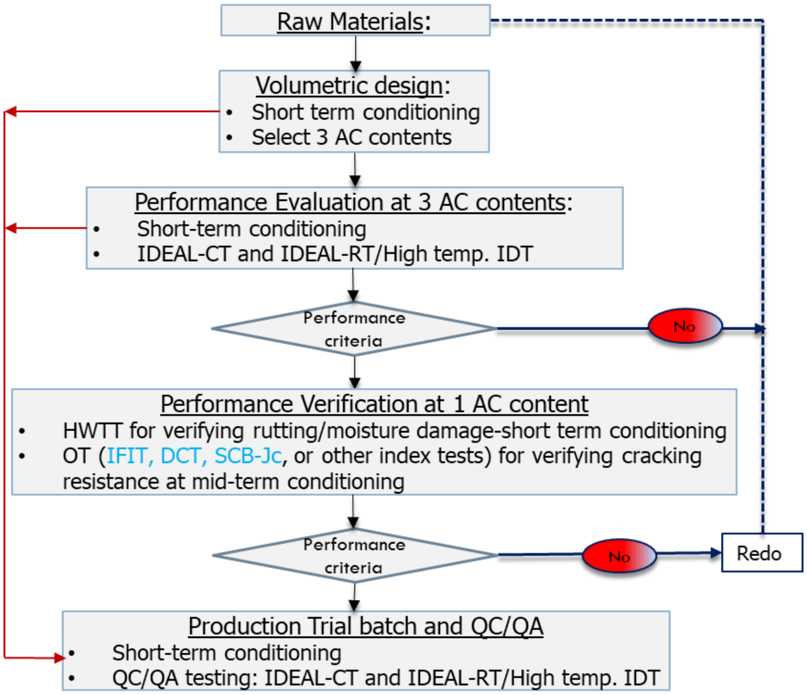


Figure 3.3 Coherent Framework for BMD/QC/QA: Example Two with Index Tests

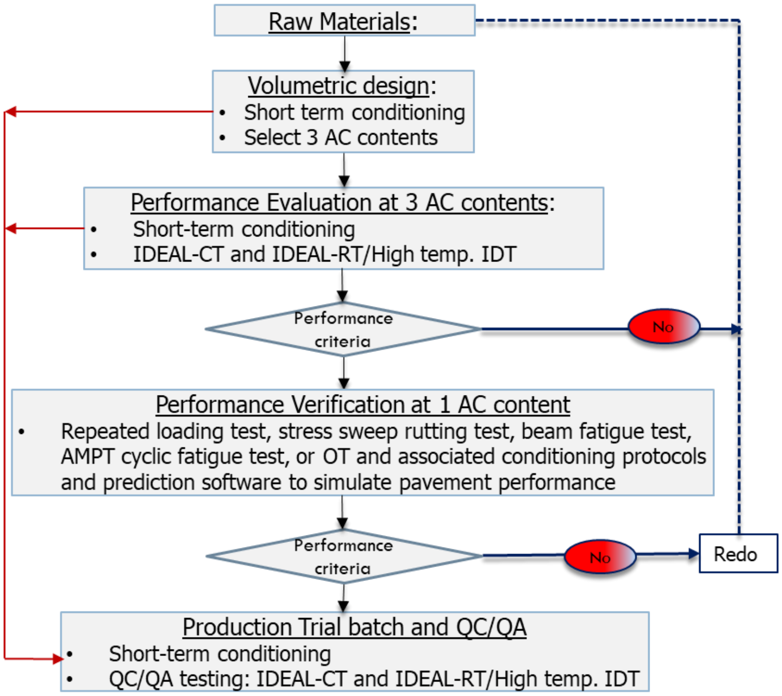


Figure 3.4 Coherent Framework for BMD/QC/QA: Example Three with Performance-Based Laboratory Tests

3.4 Laboratory aging protocols for preparing performance tests specimens

Asphalt mixtures experience two stages of the aging process in the field: (1) short-term aging at elevated high temperatures during production, transport, placement, and compaction, and (2) long-term aging under local climatic conditions. To characterize the short- and long-term aging impact, various conditioning standards or protocols have been proposed in the literature (Bell et al. 1994, Aschenbrener and Far 1994, Braham et al. 2009, Petersen 2009, Epps-Martin et al. 2014, Newcomb et al. 2015, Reinke et al. 2015, Kim et al. 2018, Chen et al. 2018). **Error! Reference source not found.** lists some major laboratory aging protocols for asphalt mixtures established since the Strategic Highway Research Program (SHRP) concluded in 1992. It is a well-known fact that the rutting and cracking properties of asphalt mixtures are not only significantly impacted by aging, but they are impacted in opposite directions, as illustrated in **Error! Reference source not found.**3.5. The longer the mixture ages, the more rutting resistance increases, but cracking resistance decreases. Thus, completely different aging protocols are needed for evaluating rutting and cracking resistance of asphalt mixtures.

Table 3.2 Major Laboratory Aging Protocols

Aging protocol		**Mix property evaluated**	**Reference**	**Note**
Short-term aging	2 hr at compaction temperature for loose mixes	Mix design: volumetric properties	AASHTO R30	
		Mix design: rutting, cracking, and moisture damage	Aschenbrener and Far (1994) Texas Department of Transportation (TxDOT) specifications (2014)	Aschenbrener and Far (1994) established it based on HWTT data
		QC: compaction density	TxDOT specifications (2014)	
	2 hr at 116°C for loose warm mix asphalt (WMA)	Mix design: mix mechanical properties	Epps-Martin et al. (2014)	This protocol was developed based mainly on resilient modulus (Mr)
	4 hr at 135°C for loose hot mix asphalt (HMA)	Mix design: mix mechanical properties	AASHTO R30 Bell et al. (1994)	This protocol was developed based mainly on Mr
	2 hr at 135°C for loose HMA	Mix design: mix mechanical properties	Newcomb et al. (2015)	This protocol was developed based mainly on Mr
Long-term aging	120 hr at 85°C for compacted specimens	Mix design: mix mechanical properties	AASHTO R30 Bell et al. (1994)	This protocol was developed based mainly on Mr
	24 to 696 hr at 95°C for loose mixes	Mix design and structural performance evaluation	Kim et al. (2018)	This protocol was developed based on asphalt binder chemistry and rheological property
	8 hr at 135°C for loose mixes	Mix design: mix mechanical properties	Chen et al. (2018)	This protocol was developed based on asphalt binder chemistry and rheological property
	24 hr at 135°C for loose mixes	Mix fracture property	Braham et al. (2009) Reinke et al. (2015)	

Note: Table 3.2 is not a comprehensive list of various aging protocols in the literature

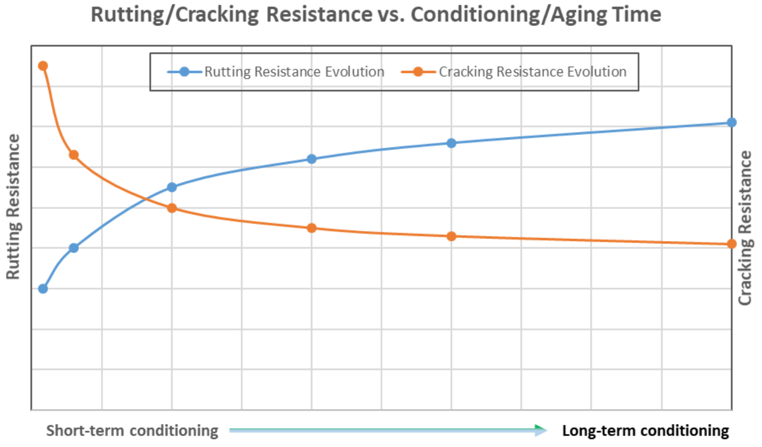


Figure 3.5 Rutting and Cracking Resistance Evolution vs. Conditioning/Aging Time

3.4.1 Aging Protocol for Preparing IDEAL-RT Specimens

As conceptualized in Figure 3.5, rutting resistance of asphalt mixtures keeps increasing with aging (or conditioning) time. Ideally, each aging condition of the full aging curve is evaluated, but it will make a routine mix design very difficult (if not impossible). The practical (or conservative) approach is to evaluate rutting resistance of an asphalt mixture at its most critical (or poorest) condition. It is well known that rutting often occurs in the very early stage of the pavement life when short-term rather than long-term aging plays a dominant effect on rutting resistance. Thus, the most critical condition for rutting is the compacted asphalt mixture immediately after the placement. The research team selected the 2-hr loose-mixture conditioning at compaction temperature for evaluating rutting resistance of asphalt mixtures during the performance evaluation and the production QC/QA testing based on the following common practices and research findings:

- The 2-hr loose-mixture aging at compaction temperature has been widely adopted by many DOTs for Superpave volumetric mix design, as specified by AASHTO R30. Furthermore, different compaction temperatures are being used by different states due to different types of binders or binder sources. To avoid any potential confusion and to ensure a smooth implementation and transition to the framework, the 2-hr loose-mix conditioning at compaction temperature is preferred.
- Many DOTs already used the 2-hr loose-mixture aging at compaction temperature for preparing rutting test specimens. For example, TxDOT uses such mixture-aging protocol for preparing HWTT specimens using Tex-241: Compacting Bituminous Specimens Using the Superpave Gyratory Compactor.
- Epps-Martin et al. (2014), under NCHRP 9-49, found that laboratory loose mixture aged for 2 hr at compaction temperature matched well with field cores taken either immediately after construction or 6 months after construction in terms of resilient modulus, Mr. Meanwhile, Epps-Martin et al. (2014) also found it difficult to define the compaction temperature for each project, so a fixed compaction temperature of 135°C was recommended for HMA mixtures and 116°C for WMA mixtures. Newcomb et al. (2015), under NCHRP 9-52, verified the aging protocol—2 hr at 135°C for HMA and 116°C for WMA—for nine field projects across the United States.

3.4.2 Aging Protocol for Preparing IDEAL-CT Specimens

In contrast to rutting, cracking frequently occurs in a later stage of pavement life. Thus, it is more appropriate to employ a long-term aging protocol for evaluating cracking resistance of asphalt mixtures. However, aging characteristics of asphalt mixtures, as illustrated in Figure 3.5, continue to evolve. Moreover, asphalt mixture aging depends on the layer depth within the pavement structure (Kim et al. 2018), and the same mixture can have completely different mixture properties if it is a surface layer than the layer 150 mm below the surface. The complexity of asphalt aging makes it extremely difficult to fully simulate the field aging in the laboratory. Further, it is often impractical to use a long-term aging protocol for conditioning loose mixtures during the production QC/QA testing. Thus, the research team devised and subsequently recommend a hybrid short- and mid-term aging protocol for preparing IDEAL-CT specimens, as noted below:

- BMD performance evaluation: IDEAL-CT specimens at multiple asphalt contents are compacted after short-term aging—the 2-hr loose mixture aged at compaction temperature.
- BMD performance verification: IDEAL-CT specimens at the balanced asphalt content are compacted after a mid-term aging period that is a 3-step process, as described below:
    - Step 1: Short-term aging process—Short-term age a loose mixture for 2 hr at its compaction temperature with the loose-mix thickness of 25–50 mm, preferably 38–50 mm, where the target depth is 38 mm +/− 12 mm. This depth of material will accommodate 4.75-mm, 9.5-mm, 12.5-mm, and 25-mm Superpave nominal maximum aggregate sizes (NMASs).
    - Step 2: Mid-term aging process—After Step 1, split the short-term aged loose mixture into multiple shallow pans (e.g., 25 mm by 330 mm by 457 mm) with a thickness of one NMAS (Kim et al. 2018); no stirring is required. Note that time and temperature used for mid-term aging of 20 hr at 100°C are intended to be the same as those used in the pressure aging vessel (PAV) conditioning for conventional and modified asphalt binders. The 20-hr aging duration is convenient for lab operation (i.e., no need of working before or after office hours) and acceptable for routine mix designs in the BMD performance verification stage. The goal of proposing a mid-term aging condition is not to simulate how aging occurs in the field but rather to stabilize the aging characteristics of loose mixtures so that it becomes possible to rank and compare cracking resistances of asphalt mixtures. More discussion is presented later when discussing acceptable criteria for the IDEAL-CT.

  - Step 3: Reheating process—After Step 2, combine the mid-term aged loose mixture into the same pan(s) used in Step 1 with the same thickness as in Step 1, then reheat the combined mixture for 2 hr at its compaction temperature and stir it each hour to ensure uniform temperature before compacting IDEAL-CT specimens.
- Production QC/QA stage: IDEAL-CT specimens are compacted after short-term aging—the loose mixture aged at compaction temperature for 2 hr.

The advantage of this hybrid aging protocol is to address the needs of evaluating the impact of both short- and mid-term aging on cracking resistance of asphalt mixtures; meanwhile, it considers the impact of aging on cracking resistance and avoids the long duration of long-term aging process in the whole BMD/QC/QA process. In this approach, the same short-term aging protocol for the IDEAL-RT is applied here to the IDEAL-CT. In such a way, the performance evaluation data can be used for developing production QC/QA acceptance criteria for both cracking and rutting resistance of asphalt mixtures.

3.4.3 Aging Protocol for Preparing HWTT Specimens

As noted previously, HWTT has been widely used for characterizing rutting and moisture damage of asphalt mixes. Aschenbrener and Far (1994) evaluated a variety of asphalt mixtures and recommended 2 hr at compaction temperature for aging the loose mixtures before compacting HWTT specimens in order to match field performance. Furthermore, many DOTs are using the 2-hr loose-mixture aging at compaction temperature as the aging protocol for preparing HWTT specimens. Thus, the 2-hr loose-mixture aging at compaction temperature is recommended for preparing HWTT specimens in the coherent BMD/QC/QA framework.

In summary, Table 3.3 lists the selected protocols for aging loose mixtures for preparing performance test specimens.

Table 3.3 Laboratory Aging Protocols for Preparing Performance Tests Specimens

Stage	Test	Loose-mix aging protocol	
BMD performance evaluation at multiple asphalt contents	IDEAL-RT	Short-term aging	2 hr at compaction temperature with a loose-mix thickness of preferably 38–50 mm; stir the loose mix every hour
	IDEAL-CT		
BMD performance verification with one asphalt content	HWTT	Short-term aging	2 hr at compaction temperature with a loose-mix thickness of preferably 38–50 mm; stir the loose mix every hour
	IDEAL-CT* (**)	Mid-term aging	(1) 2 hr at compaction temperature (just like short-term aging) (2) 20 hr at 100°C with a thickness of one NMAS and no stirring (3) 2 hr at compaction temperature (just like short-term aging)
QC/QA	IDEAL-RT	Short-term aging	2 hr at compaction temperature with a loose-mix thickness of preferably 38–50 mm; stir the loose mix every hour
	IDEAL-CT		
	Lab-molded density		

* The same long-term aging protocol for the IDEAL-CT can be used for other types of cracking tests, such as DCT, beam fatigue test, or AMPT cyclic fatigue test.

** The use of WMA additives may require aging at alternative temperatures that mimic anticipated field production temperatures in order to obtain the full benefit of producing WMA mixture.

3.5 Performance tests' criteria and strategies for meeting those criteria

Performance tests criteria are critical components of the coherent BMD/QC/QA framework. Without the performance tests criteria, one cannot determine the maximum and the minimum allowable asphalt contents to avoid premature rutting, cracking, or moisture damage problems in the field. The following subsections establish the acceptance criteria of the three performance tests.

3.5.1 IDEAL-RT Acceptance Criteria and Strategies for Meeting Such Criteria

A direct way of establishing acceptance criteria for a performance test is to construct and monitor multiple field performance test sections, but this process is very costly and takes considerable time. Alternatively, one can develop acceptance criteria for a performance test by establishing a correlation with an existing performance test and associated criteria. This chapter uses a hybrid approach for establishing the IDEAL-RT acceptance criteria, as described below.

- Development of IDEAL-RT acceptance criteria based on the relationship between IDEAL-RT and HWTT: As noted previously, the HWTT has been widely adopted by many DOTs. The most commonly used HWTT rutting parameter is total rut depth at a specified number of passes, and associated acceptance criteria have also been well established. For example, TxDOT requires a maximum rut depth of 12.5 mm at 10,000, 15,000, and 20,000 passes for asphalt mixtures with PG 64-XX, PG 70-XX, and PG 76-XX (or higher) binders, respectively. To develop a relationship between the IDEAL-RT and the HWTT, a total of 23 dense-graded mixtures (see Table 3.4) were evaluated in this study. Eleven of the 23 mixtures were laboratory-mixed and laboratory-compacted (LMLC), and the remaining 12 mixtures were reheated plant-mixed and laboratory-compacted (PMLC). For the 11 LMLC mixtures, the short-term aging protocol as described in Table 3.4 was followed to age the loose mixtures before compacting the IDEAL-RT and HWTT specimens; for the 12 PMLC mixtures, they were reheated in a forced draft oven at their respective compaction temperature. When each plant mixture became loose and workable, and reached uniformly the compaction temperature, both the IDEAL-RT and the HWTT specimens were molded. The overall heating process in the oven took approximately 2 hr. Both tests were performed at 50°C, and the test results are listed in Table 3.4.

As seen in Table 3.4, almost half of the 23 mixtures reached 12.5-mm rut depth before 20,000 passes. Thus, it becomes difficult to directly make comparisons even among 23 mixtures in terms of the rut depth at 20,000 passes. To address this issue, the research team defines a new rutting resistance index (RRI) parameter in Equation 1. The RRI parameter not only incorporates the combined effect of the number of passes and the rutting depth, it also addresses the nonlinear impact of number of loading passes (or repetitions) on pavement rutting (or permanent deformation) through using N^0.3 rather than N^1, unlike the work by Wen et al. (2016). The calculated RRI for each mix is listed in Table 3.4 as well.

$$RRI = N^{0.3}\left(1 - \frac{RD}{25.4}\right) \quad (1)$$

where,

RRI = Rutting resistance index.

N = 20,000 or number of passes reaching 12.5-mm rut depth.

RD = Rut depth at 20,000 passes or 12.5 mm for those reaching 12.5- mm rut depth before 20,000 passes.

The target maximum RRI values for mixtures with PG64-XX, PG70-XX, and PG76-XX are 8, 9, and 10, respectively.

The relationship between HWTT (RRI) and IDEAL-RT (RTIndex) is shown in Figure 3.6. The 95 and 98 percent confidence intervals are also added to the graph. As seen in Figure 3.6, there is a good linear relationship between RRI and RTIndex. The larger the RTIndex, the larger the RRI and the better rutting resistance.

Based on the relationship shown in Figure 3.6 and the existing HWTT acceptance criteria, one can calculate the RTIndex values corresponding to HWTT rutting criteria for mixtures with PG64-XX, PG70-XX, and PG76-XX. Note that the research team used different confidence intervals when setting the minimum RTIndex requirements for different mixtures, in consideration of the importance of avoiding rutting and potential safety issues, as listed below:

  - For mixtures with PG64-XX (or lower) with 95 percent confidence: RTIndex ≥ 60.
  - For mixtures with PG70-XX with 95 percent confidence: RTIndex ≥ 65.
  - For mixtures with PG76-XX (or higher) with 98 percent confidence: RTIndex ≥ 75.

- Generally, it is expected that an asphalt mixture with PG64-XX used in a hot climate ruts quickly under heavy traffic; in contrast, an asphalt mixture with PG76-XX will have no (or much less) rutting problem when used in the same environment. Alternatively, a mixture with RTIndex < 60 will very likely rut prematurely in comparison to a mixture with RTIndex ≥ 75, which is discussed in the next subsection.

Table 3.4 IDEAL-RT and HWTT Test Results of 23 Asphalt Mixes

Asphalt mix					HWTT			IDEAL-RT
Mix No.	Virgin asphalt binder	RAP by weight of total mix (%)	RAS by weight of total mix (%)	Total binder content (%)	Rut depth at 20,000 passes (mm)	No. of passes at 12.5-mm rut depth	RRI	RT_{Index}
1	PG58-28	10	0	5.4	N/A	4,860	6.5	35.2
2	PG64-28	20	0	5.3	N/A	18,792	9.7	58.3
3	PG64-22	15	2	4.8	6.82	N/A	14.3	90.6
4	PG70-22	0	0	4.8	N/A	12,864	8.7	61.0
5	PG64-28	15	0	4.8	N/A	15,004	9.1	59.8
6	PG64-22	15	2	5.3	9.63	N/A	12.1	66.9
7	PG64-28	20	0	5.7	N/A	12,652	8.6	67.7
8	PG64-28	20	0	6.4	N/A	18,980	9.8	58.2
9	PG58-28	7	3	6.0	N/A	10,196	8.1	64.5
10	PG64-22	15	5	5.0	3.21	N/A	17.0	107.5
11	PG64-22	30	0		N/A	12,856	8.7	64.3
12	PG64-22	40	0		4.15	N/A	16.3	97.1
13	PG64-22	20	0	5.1	N/A	13,056	8.7	72.0
14	PG76-22	0	0	6.8	11.51	N/A	10.7	78.4
15	PG64-22	25	0	5.3	10.59	N/A	11.4	62.6
16	PG76-22	0	0	5.5	2.89	N/A	17.3	118.1
17	PG64-22 and 3.2% rejuvenator	25	0	5.0	12.5	7,696	7.4	53.4
18	PG64-22 and 4.0% rejuvenator	25	0	5.0	12.5	7,376	7.3	43.1
19	PG64-22	15	2	5.0	8.94	N/A	12.6	74.9
20	PG64-22	0	0		12	N/A	10.3	80.7
21	PG64-22	20		5.2	4	N/A	16.4	92.7
22	PG64-22	10	5	5.2	1.85	N/A	18.1	127.1
23	PG70-22	10	2	5.6	2.28	N/A	17.8	119.8

Note: N/A stands for not applicable.

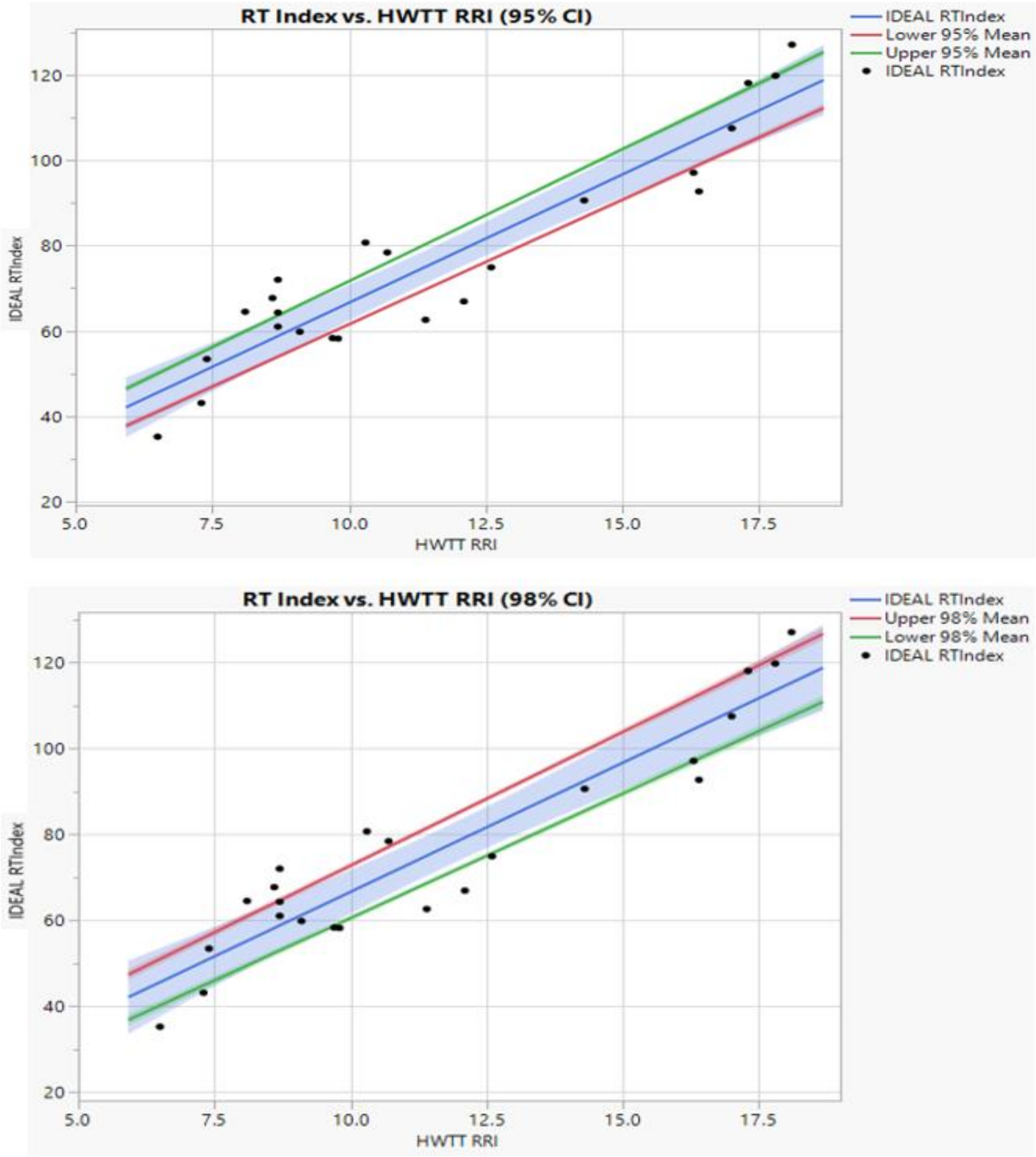


Figure 3.6 Linear Relationship between IDEAL-RT and HWTT

- Strategies for meeting the IDEAL-RT criteria: In the last two decades, rutting has been substantially minimized, if not completely eliminated, from asphalt pavements. To improve rutting resistance and increase RTIndex, one can employ the following one or combined measures:
    - Use stiffer binders, typically polymer-modified binders.
    - Replace natural sand with manufactured sand or crushed particles.

- Use angular (or more crushed) aggregates.
- Reduce asphalt content.
- Use RAP/RAS (if allowed by the agency).
- Change binder source (Note that the asphalt binders with the same PG do not always perform the same).

3.5.2 IDEAL-CT Acceptance Criteria and Strategies for Meeting Such Criteria

As discussed previously, cracking resistances of asphalt mixtures are evaluated under two aging conditions: short-and mid-term aging. Thus, two IDEAL-CT criteria are established below.

- Short-term aging IDEAL-CT criteria: In a previous study, Zhou et al. (2020) established IDEAL-CT criteria for dense-graded mixtures and stone matrix asphalt (SMA) mixtures (Table 3.5) based on the correlation between the IDEAL-CT and Texas OT. Note that the CTIndex criteria in Table 3.5 are intended to address reflective cracking in asphalt overlays because reflective cracking in overlays is the primary distress pavement each DOT is facing. Furthermore, Zhou et al. (2020) showed that CTIndex = 90 corresponds to flexibility index (FI) = 8 based on limited data. A similar relation was reported by Romero (2021)

Table 3.5 IDEAL-CT Acceptance Criteria for Different Mixtures

Mix type	**SMA**	**Superpave dense-graded mixes**
Minimum CT_{Index}	135	90

- Validation of the short-term aging CTIndex criteria: To validate the short-term aging CTIndex criteria, the research team turned to LTPP-SPS10, Warm Mix Asphalt (WMA) Overlay of Asphalt Pavements. Six test sections were constructed on SH66, west of Yukon, Oklahoma, in November 2015. Before the 50-mm asphalt overlay, LTPP surveyed and recorded existing pavement distresses of the six test sections. All test sections exhibited a large amount of cracking, except Section 400A62, with no transverse cracking. For the purpose of validating the short-term aging CTIndex criteria with a focus of reflective cracking, Section 400A62 was excluded from this study. Thus, only five test sections (400A01, 400A02, 400A03, 400A61, and 400A63) are employed here for the validation of the CTIndex criteria. The latest distress survey data in LTPP-Inforpave database were recorded in March 2019 after 40 months trafficking (Figure 3.7). Note that the reflective cracking rate is the ratio of the transverse cracking length observed on March 5, 2019, to the transverse cracking length observed on April 15, 2015 (the last survey before the asphalt overlay). Meanwhile, plant mixture from each test section was tested under the IDEAL-CT by following ASTM D8225. For each test section, three replicates of IDEAL-CT specimens with 7±0.5 percent air voids were molded after short-term aging. The IDEAL-CT was performed at 25°C with a loading rate of 50 mm/min. Figure 3.7 shows both the IDEAL-CT test results and the reflective cracking of each test section. As seen in Figure 3.7, the CTIndex has a very good correlation with field reflective cracking development; the smaller the CTIndex value, the higher the reflective cracking rate. Furthermore, when the CTIndex value is larger than 90, its reflective cracking rate is less than 10 percent after 40 months trafficking for a 2-inch asphalt overlay. Thus, CTIndex = 90 is a reasonable number to start with, although a more robust field validation effort is still needed.

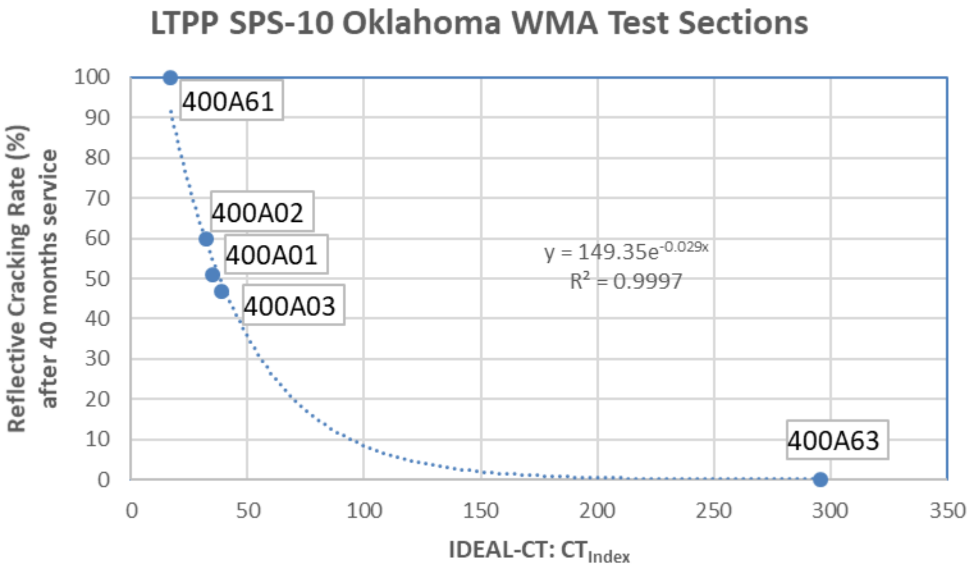


Figure 3.7 Validation of the Short-Term Aging IDEAL-CT Criteria

- Mid-term aging IDEAL-CT criteria: Since the mid-term aging concept has not been proposed until now, no information is available in the literature. Instead of directly comparing the mid-term aging to field aging, this chatper establishes the mid-term aging IDEAL-CT criteria by evaluating CTIndex reduction from the CTIndex value at the short-term aging (2 hr at compaction temperature). If a common CTIndex reduction exists, the mid-term aging IDEAL-

CT criteria can be developed using the short-term aging IDEAL-CT criteria (CTIndex ≥ 90) × the common CTIndex reduction. Based on this philosophy, six typical asphalt mixtures often used in Texas were employed to evaluate CTIndex evolution and reduction with aging. Table 3.6 details each mixture information. For each mixture, the loose mix was aged under four different aging conditions:

- 2hr at compaction temperature (short-term aging protocol proposed here).
- 4hr at 135°C (AASHTO R 30 aging protocol).
- 2 hr at compaction temperature + 20 hr at 100°C + 2 hr at compaction temperature (mid-term aging protocol proposed here).
- 2 hr at compaction temperature + 144 hr at 95°C + 2 hr at compaction temperature (long-term aging protocol equivalent to asphalt binder 40 hr PAV aging [Kim et al. 2018]).

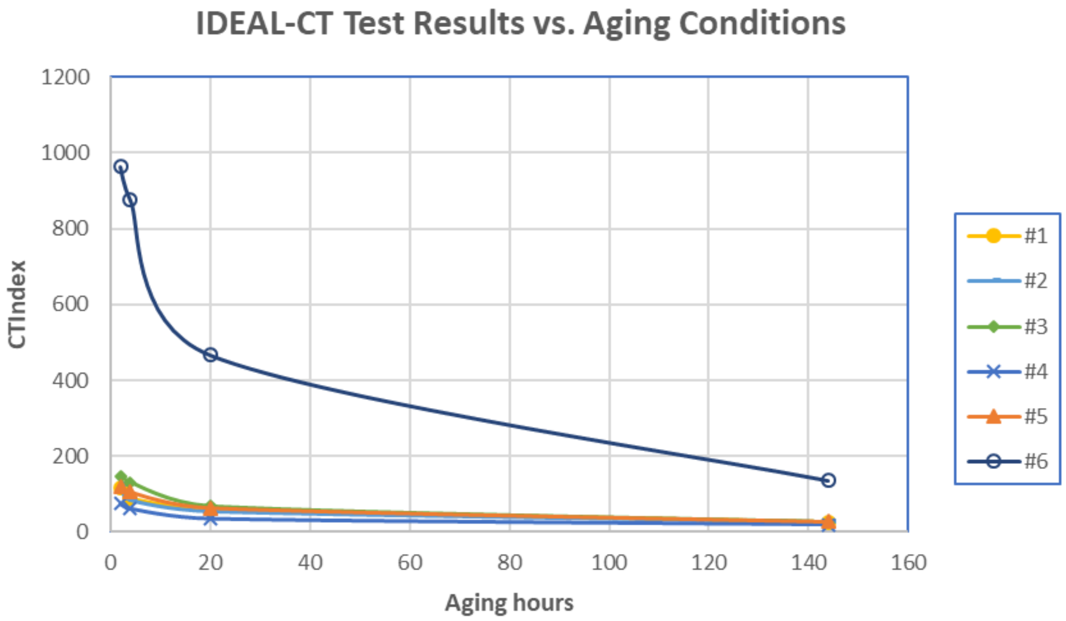
  After the aging, three IDEAL-CT specimens were compacted using the Superpave gyratory compactor and then were tested by following ASTM D8225. The average of the test results for each mixture under four aging conditions are shown in Figure 3.8(a). In order to investigate the CTIndex reduction with aging, the measured CTIndex values at different aging conditions were normalized to the CTIndex value at the short-term aging (2 hr at compaction temperature), as shown in Figure 3.8(b).

It can be seen from Figure 3.8(b) that the CTIndex values of all six mixtures consistently drop with increased aging time, although the dropping amount of the CTIndex value is different for each mixture. A simple statistical analysis was performed to determine the average drop and associated standard deviation at the mid- and long-term aging conditions. As shown in Figure 3.8(b), the CTIndex drop and standard deviation are 0.48 and 0.03 for the mid-term aging and 0.21 and 0.04 for the long-term aging, respectively. After considering the variation of the CTIndex drops, the research team recommended using 0.45 (= 0.48–0.03) as the CTIndex drop to calculate the minimum requirement for the CTIndex after the mid-term aging. Consequently, the research team recommended the following mid-term aging IDEAL-CT criteria: CTIndex ≥ 40 (= 90*0.45).

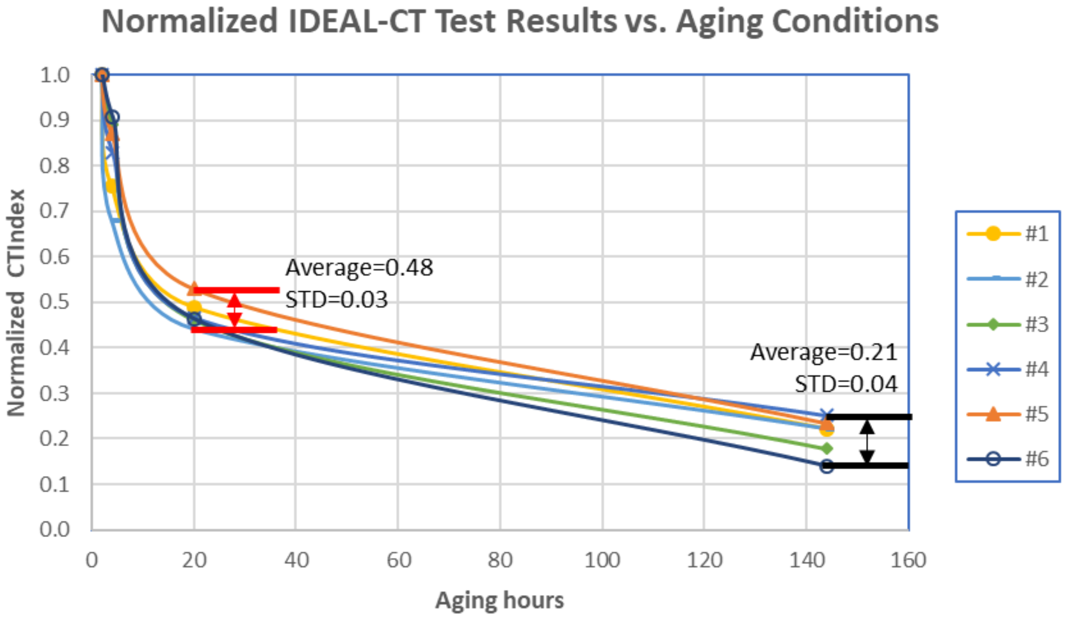
Table 3.6 Six Mixtures Used for Developing Mid-term Aging IDEAL-CT Criteria

Mixture type		Virgin binder	RAP/RAS (%)	Rejuvenator	Total asphalt content (%)
#1	12.5-mm Superpave	PG70-22	10% RAP	None	5.3
#2	12.5-mm Superpave	PG70-22	None	None	5.6
#3	12.5-mm Superpave	PG64-22	15%RAP/2%RAS	Bio-rejuvenator	5.2
#4	9.5-mm Superpave	PG76-22	None	None	5

#5	9.5-mm Superpave	PG76-22	None	None	5.5
#6	12.5-mm SMA	PG76-22	None	None	6.3



(a) CT_{Index} vs. Aging



(b) Normalized CT_{Index} vs. Aging

Figure 3.8 IDEAL-CT: Cracking Resistance Evolutions with Aging

- Strategies for meeting the IDEAL-CT criteria: Asphalt industry and pavement engineers have been addressing the cracking or durability problem for the last two decades. Different approaches have been tried in both the laboratory or in the field (Zhou et al. 2007, Im et al. 2015, West et al. 2018, Mogawer et al. 2019, Zhou et al. 2011, Zhou et al. 2013, Zhou et al. 2020, Karki and Zhou 2018, Tran et al. 2019). To improve cracking resistance and increase the CTIndex, one can employ the following one or combined measures:
    - Increase asphalt content by increasing design density (such as regression air void approach), varying aggregate gradation away from the maximum density line, and reducing design gyration (in case of keeping the same aggregate gradation).
    - Use polymer-modified binders designed for cracking resistance.
    - Reduce the RAP/RAS amount.
    - Use rejuvenators.
    - Use chemical WMA and lower production temperatures.
    - Add fibers.
    - Use less absorptive aggregates.
    - Change binder source (Note that the asphalt binders with same PG do not always perform the same).

Multiple field test sections have been constructed and monitored in Texas. Karki and Zhou (2008) reported that effective strategies for improving cracking resistance are to increase asphalt binder content and then use polymer-modified binders designed for cracking resistance (such as polymer-modified softer binders, PGxx-28 or PGxx-34).

3.5.3 HWTT Acceptance Criteria and Strategies for Meeting Those Criteria

As stated previously, the HWTT and associated criteria have been well established. Rut depth at a specified number of passes is the most commonly used parameter for evaluating rutting and moisture damage potential, although some DOTs have an additional requirement of stripping inflection point. To be consistent with current practices, the research team adopted exactly the same criteria being used by DOTs. These same criteria will be used in the framework to exclude any potential moisture susceptible mixtures.

Aggregate characteristics (such as minerals, surface chemistry, porosity) have significant influence on moisture susceptibility of asphalt mixtures. To address moisture damage of asphalt mixtures, the two most often used approaches are chemical liquid antistripping agents and lime (Hicks 1991).

3.6 Case study: demonstration of the coherent BMD/QC/QA framework

The purpose of the case study is to demonstrate the developed the framework with three performance tests selected previously: IDEAL-CT, IDEAL-RT, and HWTT. The asphalt mixture used for this demonstration is an actual mixture designed and placed for an accelerated pavement testing project in Texas. The BMD mixture is a 12.5-mm Superpave mixture with limestone aggregates, a virgin binder PG64-22, and a binder replacement of 24 percent from RAP. **Error! Reference source not found.** shows the aggregate gradation, and the same gradation is used for this whole demonstration process. A step-by-step design process is described below.

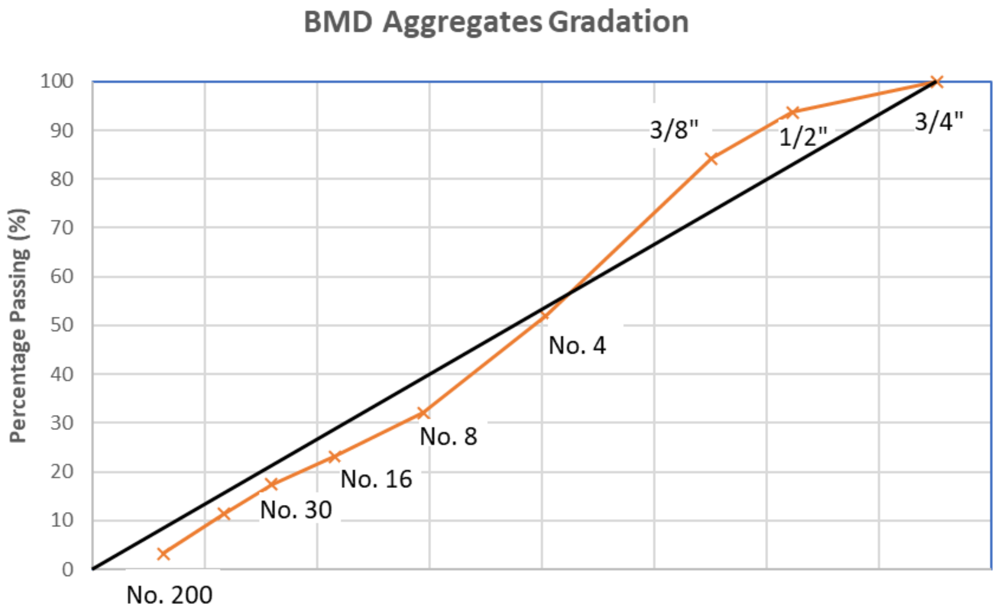


Figure 3.9 Aggregates Gradation of the 9.5-mm Dense-Graded Mix

3.6.1 Volumetric mix design

The original mix design calls for a 12.5-mm Superpave virgin mixture with PG70-22. According to TxDOT's 2014 construction specification, the virgin binder PG70-22 can be replaced with a PG64-22 binder when 15 percent RAP or more is used in the asphalt mixture. However, the temperatures for mixing the aggregates and asphalt binder and aging the loose mixture should be the same temperatures as used for the original asphalt binder. In this case, the mixing and aging temperatures for a PG70-22 binder are 149°C and 135°C, respectively. Thus, the same mixing and aging temperatures were used for designing the 12.5-mm Superpave virgin mixture with PG64-22 binder and a 25 percent RAP (in weight of the total mixture). The number of gyrations selected for this case study was Ndesign = 35, which is a common number for Superpave mixtures with limestone aggregates. Following the aging protocols established previously, the loose mixtures at three asphalt binder contents—4.5, 5.0, and 5.5 percent—were aged for 2 hr at 135°C (aging temperature, or compaction temperature). The measured densities of the compacted asphalt mixture at three asphalt contents are shown in Figure 3.10.

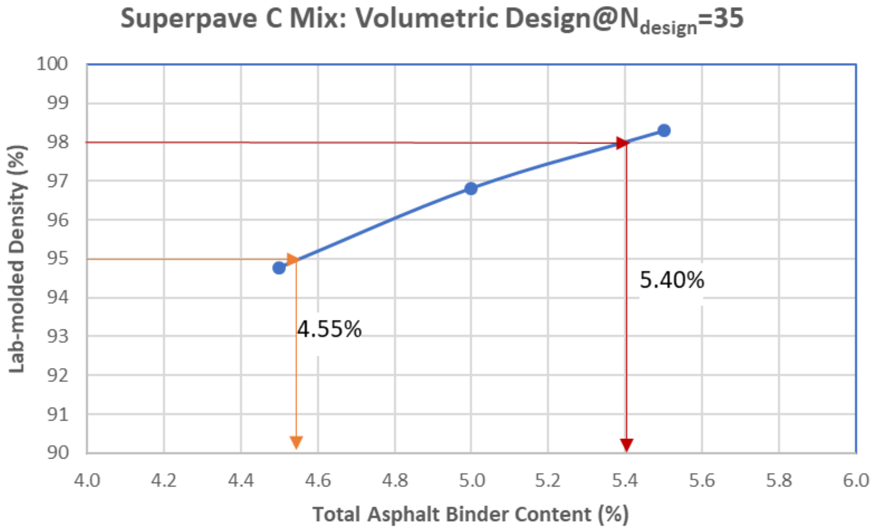


Figure 3.10 Volumetric Mix Design: Density vs. Asphalt Content

As discussed previously, three asphalt binder contents within the density range of 95 to 98 percent were selected for performance evaluation. For this case, the asphalt contents corresponding to 95 percent and 98 percent density were 4.6 percent and 5.4 percent, respectively, as shown in Figure 3.10. Within such range, the research team selected three asphalt contents—4.7, 5.0, and 5.3 percent—for performance evaluation in the next step.

3.6.2 Performance Evaluation of Asphalt Mixtures with Three Asphalt Contents

The 12.5-mm Superpave mixture with PG64-22 and 25 percent RAP (in weight of the total mixture) was mixed with three asphalt binder contents: 4.7, 5.0, and 5.3 percent. The asphalt binder content is the total asphalt binder content, including the binder from the RAP material. Then, the three loose mixtures were aged at 135°C for 2 hr before compacting the IDEAL-CT and -RT specimens. For each asphalt content, a total of six replicates of specimens at the air voids of 7±0.5 percent were molded, three specimens for the IDEAL-CT test and another three specimens for the IDEAL-RT test. Figure 3.11 shows the IDEAL-CT and -RT test results.

Based on the IDEAL-CT and -RT criteria—CTIndex ≥ 90 and RTIndex ≥ 65 for asphalt mixtures with PG70-22—the research team established a balanced zone, as shown in Figure 3.11. For this case, the boundary asphalt binder contents of the balanced zone were 5.10 to 5.24 percent, within which the asphalt mixture met both rutting and cracking criteria. Considering the fact that the asphalt mixtures often become either dryer or finer, the selection of the asphalt content favored the upper boundary of the balanced zone. For this case, the total asphalt binder content of 5.2 percent was selected for performance verification in the next step.

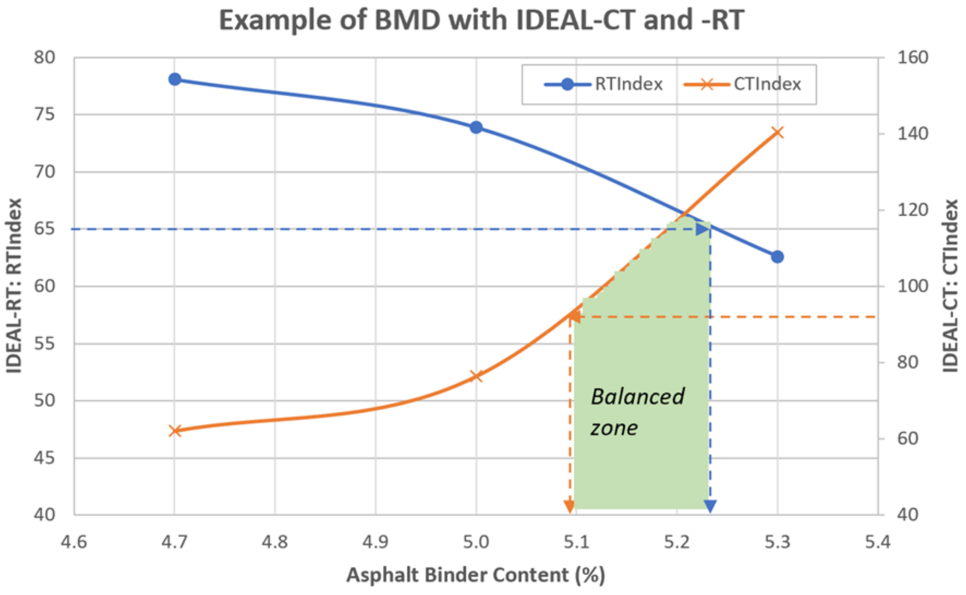


Figure 3.11 Performance Evaluation with Three Asphalt Binder Contents

3.6.3 Performance Verification with the Selected Asphalt Binder Content

For this case study, the asphalt mixture was paved in Dallas District, Texas. Low-temperature cracking was not a major concern. Thus, the research team verified three distresses only: cracking, rutting, and moisture susceptibility of the 12.5-mm Superpave mixture with PG64-22 and 25 percent RAP (in weight) at the total asphalt binder content of 5.2 percent. As noted previously, the IDEAL-CT is employed for evaluating cracking resistance of the mixture after mid-term aging, and the HWTT is used for rutting and moisture susceptibility of the mixture after short-term aging. One set of the HWTT specimens were compacted using the Superpave gyratory compactor after short-term aging the loose mix for 2 hr at 135°C. For the IDEAL-CT, three test specimens were prepared using the Superpave gyratory compactor after mid-term aging the loose mixture, as detailed in Table 3.3.

The HWTT was performed at 50°C, following Tex-242-F: Hamburg Wheel Tracking Test, and the test result is shown in Figure 3.12. For this case study, since the mix was originally designed with PG70-22 binder, the associated requirement of the rut depth at 15,000 passes was less than 12.5 mm. As shown in Figure 3.12, the rut depth at 15,000 passes was 7.5 mm, and it met TxDOT's requirement for both rutting and moisture susceptibility.

The mid-term aged IDEAL-CT specimens were tested at 25°C—following ASTM D8225—and the average CTIndex values of the mid-term aged specimens was 55. Compared to the CTIndex = 117 (estimated from Figure 3.11) after the short-term aging, the CTIndex ratio of the mid-term aging to the short-term aging was 0.47 (55/117 = 0.47), which is larger than the minimum acceptance value: 0.45. Thus, the 12.5-mm Superpave mixture with PG64-22 and 25 percent RAP (in weight) at the total asphalt binder content of 5.2 percent passes the cracking requirement after the mid-term aging. Thus, the next step was to develop the QC/QA testing plan and acceptance criteria for plant production.

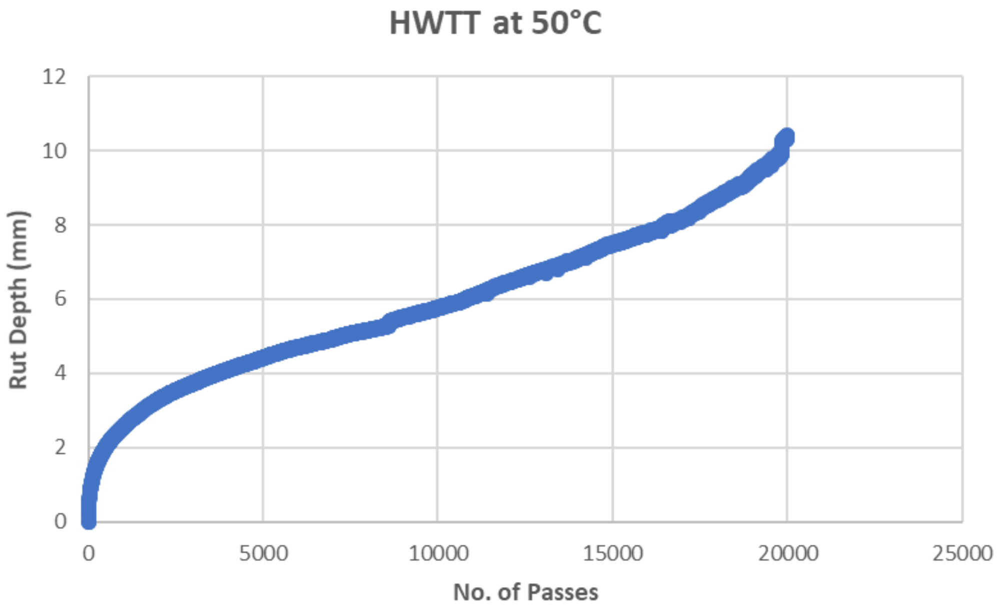


Figure 3.12 HWTT Test Results of the 12.5-mm Superpave Mixture at the Asphalt Content of 5.2 Percent

3.6.4 Development of QC/QA Testing Plan and Acceptance Criteria

The new QC/QA testing plan focuses on three material properties: compacted density, CTIndex, and RTIndex corresponding to the balanced asphalt content. For this case study, the balanced asphalt content was 5.2 percent. As seen in Figure 3.10, 5.2 percent asphalt content corresponds to 97.5 percent density. As discussed previously, the maximum density cannot be greater than 98 percent. Thus, considering variability of plant production, the QC/QA density requirement will be 97.5±0.5 percent.

For the CTIndex and RTIndex acceptances, two ways to establish the acceptance criteria for the IDEAL-CT and -RT exist, which are described below:

- General minimum value approach: QC/QA acceptance for cracking and rutting will be the same minimum cracking and rutting requirements as those used for the mix design. For this case study, the minimum cracking and rutting requirements were CTIndex ≥ 90 and RTIndex ≥ 65. As long as the measured CTIndex and RTIndex values are equal or larger than these criteria, the mix passes the production QC/QA check, regardless of the CTIndex and RTIndex values of the original mix design corresponding to the balanced asphalt content. This approach is clear, straightforward, and easy to understand and implement, but it may lower the quality of an asphalt mix. For example, during the mix design, consider an asphalt mix that has a very good cracking resistance of CTIndex = 180. If the minimum value approach is employed during the production QC/QA testing, the cracking resistance of the production mix may drop to CTIndex = 90, but it still passes the production QC/QA testing because it meets the minimum cracking resistance requirement, although the mix cracking resistance is significantly reduced during plant production.

- Mix-specific acceptance approach: Contrary to the general minimum value approach, the mix-specific approach establishes mix-specific acceptance criteria for the IDEAL-CT and -RT at the balanced asphalt content during the mix design stage. The acceptance criteria are still minimum values of CTIndex and RTIndex at the balanced asphalt content. For this case study, the minimum CTIndex and RTIndex values were determined from the data presented in Figure 3.11; they are CTIndex ≥117 and RTIndex ≥ 67.
- For this case, both the mix-specific and the general acceptance criteria of CTIndex and RTIndex were close. Furthermore, considering production variability, the chosen acceptance criteria of CTIndex and RTIndex for QC/QA testing were as follows: CTIndex ≥ 90 and RTIndex ≥ 65 for this case study. Note that the original mixture was a virgin PG70-22 binder. Thus, RTIndex ≥ 65 was selected.

3.6.5 Production QC/QA Testing

The asphalt mixture designed above was one of seven mixtures evaluated under a large, accelerated pavement testing study. This mixture was actually produced at an asphalt plant located in Dallas, Texas, and then paved on the accelerated pavement testing site. During the production, the asphalt mixture was sampled at the asphalt plant just like the conventional QC sampling. The sampled asphalt mixture was conditioned at the plant QC lab oven for 2 hr at the compaction temperature of 135°C before compacting specimens for density measurement (at Ndesign = 35) and then for three IDEAL-CT specimens at 7±0.5 percent air voids and another three IDEAL-RT specimens at 7±0.5 percent air voids. The QC test results are listed in Table 3.7. Thus, the produced asphalt mixture at the asphalt plant met all the acceptance criteria.

Table 3.7 QC/QA Requirements and Actual Test Results

QC/QA Parameters	Compacted Density (%)	CT_{Index}	RT_{Index}
Acceptance criteria	97.5±0.5	90	65
Actual QC test result	97.5	99	75

3.7 Summary

Asphalt mixtures are becoming increasingly complex, and given the ever-changing components of asphalt mixtures, both BMD and production QC/QA are critical in order for a mixture to perform well in the field. This chapter established a coherent BMD/QC/QA framework that has four interconnected components: (a) volumetric mix design and selection of multiple asphalt contents for mixture performance evaluation, (b) mixture performance evaluation at multiple asphalt contents and selection of the balanced asphalt content, (c) mixture performance verification at the balanced asphalt content, and (d) production QC/QA testing. One set of laboratory tests and associated acceptance criteria are recommended for the mixture performance evaluation and the production QC/QA testing and the other set is recommended for the mixture performance verification wherein DOTs can choose their preferred performance verification tests. Furthermore, two practical loose-mixture aging protocols were developed, one for short-term aging used in the process of the volumetric mix design, the mixture performance evaluation, and the production QC/QA testing, and the other for mid-term aging employed in the mixture performance verification. The recommended short-term aging protocol is to age the loose mixture in a force draft oven for 2 hr at the mixture compaction temperature, while the mid-term aging protocol consists of three steps: (1) short-term aging, (2) 20-hr loose-mixture aging at 100°C, and (3) reheating for compaction. A case study was presented in this chapter to demonstrate the whole process of the framework, including the actual plant production QC testing.

Chapter 4 Study of variability of tests at intermediate temperature

4.1 Overview

Testing of asphalt mixtures must consider the variability of the results. If the variability in mix performance tests is too high, then the results are problematic. Even if a test can be shown to follow an expected trend that might relate to performance, the prediction can be lost if both precision and accuracy are in doubt. A measurement is considered precise if all results are close to the same value; precision of a test can be quantified with parameters such as standard error or coefficient of variation. Of course, precision alone is of little value if the actual property value is not measured (i.e., the measurement is not accurate). Unfortunately, given the nature of material testing, accuracy is not easily quantified and practitioners often look at casual trends as a way to determine if the test is accurate.

The coefficient of variation, defined as the standard deviation divided by the mean and stated as a percent, can be used to quantitatively evaluate the precision of the test and, more specifically, evaluate the repeatability of a test in a single lab with a single operator. Based on experience with asphalt mixtures testing, a coefficient of variation less than 25% is desirable to assure that the results are meaningful for acceptance testing.

To evaluate between-laboratory variability, a more sophisticated method must be employed. Since three laboratories contributed to the results, it is necessary to look at both the variation between labs as well as within labs. A single variable ANOVA evaluation will help to see whether all of the data comes from the same population. To make this analysis meaningful, all factors except the laboratory were kept as close to each other as possible. For this evaluation, the calculated F value must be smaller than the F-crit. value at the 0.05 confidence interval to conclude that the labs are getting the same answer.

4.2 Mixture design

Two asphalt mixtures were used, the mixtures are representative of material used by Utah DOT but were modified to eliminate RAP as it was felt that RAP would introduce an additional variable.

Two aggregate gradations and two asphalt binders were selected for this study. Aggregate gradations for mixes A and B are shown in Table 4.1, and mix properties are shown in Table 4.2.

Table 4.1 Aggregate Gradation

	Mix Aggregate Gradations Percent Passing							
	19mm	12.5 mm	9.5 mm	4.75 mm	2.36 mm	1.8mm	0.30mm	0.075mm
Mix	3/4	1/2	3/8	#4	#8	#16	#50	#200
A	100	89.0	80.0	48.0	28.0	17.0	10.0	7.1
B	100	97.0	87.0	45.0	28.0	21.0	12.0	5.8
1% Lime								

Table 4.2 Mixture Characteristics

Mix Designation	**Geology Description**	**Aggregate Bulk Specific Gravity, G_{sb}**	**Aggregate NMAS, mm**	**Design Gyrations**	**Binder Content, %**
A	Hard Limestone	2.679	19.5	100	1.6 (m) 12.3 (v)
B	Quartzite and Granite	2.668	12.5	75	5.3 (m) 14.1 (v)

(m) by mass, (v) by volume

Laboratory samples were prepared in accordance with AASHTO TP-124 and ASTM D 8225 and mixed and compacted at the temperatures designated by the binder manufacturer. The mix was aged at compaction temperature for two hours, which is the standard practice at UDOT for all performance testing. The samples were compacted to height to achieve target air voids of 7.5 ±0.5%. Three different laboratories were involved in the study: UDOT Central Lab, PEPG, and University of Utah.

Samples for the I-FIT were tested in a previous study (Romero and VanFrank, 2019), samples for the IDEAL-CT study are described in Table 4.3. A full test set consisted of three sets of three pucks, totaling 9 pucks per lab per condition.

Table 4.3 Study Matrix for IDEAL-CT testing

				Lab		
Test	Configuration	Mix	Binder	UDOT	PEPG	UofU
IDEAL-CT	115mm cut to 62 mm	A	64-34	X	X	X
		B		X	X	X
		A	70-28	X		
		B		X	X	X

	62mm uncut	A	64-34	X		
		B		X	X	X
		B	70-28	X	X	X
	75mm uncut	B	64-34	X	1/3	1/3

4.3 SCB flexibility index variability I-FIT

The Flexibility Index based on the Illinois version of the SCB test was the subject of a previous study (Romero and VanFrank, 2019). In that study, the within-lab variability and the between-lab reproducibility were investigated. Based on statistical evaluations, the following was found:

- The study showed that even though there is a slight skewness towards the high end, the results from the test can be considered normally distributed, thus descriptive statistics can be used.
- The comparison of results between 3 labs indicated that, while the results are reproducible within each lab on repeated days, it is possible that a bias is introduced by a lab. Thus, it is important to verify on a regular basis that all labs are getting statistically similar results.
- The study revealed that at least 8 samples should be tested to obtain an average that represents the actual value within 20%. This requires compaction of 2 gyratory pucks.
- A coefficient of variation (CoV) between 20% - 30% was observed for samples cut from one puck. When comparing the average of four samples cut from one puck to another similar puck, the CoV was around 11%.
- There was little advantage found in performing the tests at a slower loading rate; however, more testing is recommended.

4.4 IDEAL-CT index variability

4.4.1 IDEAL-CT results

To determine whether the test results from the I-FIT and IDEAL-CT tests were similar, Coefficient of Variation (CoV) for within-lab and ANOVA "F" values needed to be determined. Two aggregate blends and two binder grades were tested. Although the I-FIT binders were the same grade, the sources were different. The results are tabulated in Table 4.4. Note that not all mixes were tested in all labs. Analysis of the data indicates that, even though it has been shown that these fracture data are generally normally distributed, there is a tendency to exhibit single large values (i.e., high value outliers). About half of the time, this is the case with these results.

During testing it was observed that 62-mm, uncut samples, required high numbers of gyrations to compact to height and target density. This condition prompted the researchers to add a 75-mm tall puck to the Mix B sample set to determine if compaction was easier. The trial was successful with compaction not exceeding 60 gyrations.

Table 4.4 IDEAL-CT Test Results

IDEAL CT Index Results								
	Lab			Lab			Lab	
UDOT	PEPG	UofU	UDOT	PEPG	UofU	UDOT	PEPG	UofU
Mix A PG 64-34 62mm Cut			Mix A PG 64-34 62mm Uncut					
161.9	_146.7*_	93.8	173.0		93.8			
137.8	321.2	131.3	198.1		131.3			
150.8	285.3	132.8	127.5		132.8			
164.3	_382.4*_	134.6	165.1		134.6			
110.5	306.6	146.0	154.6		146.0			
122.8	283.1	133.7	157.8		133.7			
147.5	298.6	181.1	77.8		181.1			
167.4	341.3	149.4	104.4		149.9			
170.3	316.4		94.7		276.2			
Mix A PG 70-28 62mm Cut								
199.1								
185.7								
203.2								
150.0								
133.5			_* values are more than two standard deviations from the mean_					
184.6								
171.0								
165.0								
161.4								
Mix B PG 64-34 62mm Cut			Mix B PG 64-34 62mm Uncut			Mix B PG 64-34 75mm Uncut		
531.7	515.6	280.9	173.0	274.5	227.8	439.7	404.3	305.1
320.0	464.4	464.6	198.1	337.3	_556.2*_	371.5	264.7	373.8
377.9	395.8	319.0	127.5	284.9	396.1	488.2	320.0	433.2
485.2	375.8	310.8	165.1	289.5	203.9	373.3		520.2
265.1	255.8	385.9	154.6	92.0	379.2	_767.8*_		456.5
316.9	305.6	518.4	157.8	491.6	331.1	447.7		558.0
309.0	484.6	285.9	77.8	526.4	253.5	330.0		362.9
_937.3*_	368.2	451.1	104.4	377.2	337.1	389.1		354.9
_694.1*_	437.1	529.5	94.7	596.7	382.2			393.8
217.6								
321.5								
343.7								
Mix B PG 70-28 62mm Cut			Mix B PG 70-28 62mm Uncut					
412.7	685.4	466.5	368.9	843.9	378.7			
409.8	1142.2	368.6	675.9	1181.1	429.6			
420.7	904.6	435.0	369.2	699.0	566.9			
464.5	1101.8	682.3	374.6	1209.4	529.7			
618.0	718.2	773.4	481.0	168.3	490.8			
491.4	880.7	618.3	397.2	249.3	605.1			
414.2	547.6	471.3	395.7	264.7	573.6			
457.3	1479.0	399.8	591.3	608.0	469.4			
639.8	781.3	591.9	589.1	665.4	916.3			

4.4.2 Evaluation of IDEAL-CT results

4.4.2.1 Cracking Test Normal Distribution

In previous research, it was noticed that although the fracture index, FI, results are normally distributed, there is a tendency toward single-incidence, high values. This is also noted in this data. To obtain reasonably repeatable results, the researchers discarded values falling two standard deviations above the mean. A comparison of CoV values before and after this modification are given in Table 4.5. As expected, the CoV is greatly reduced by the adjustment.

Table 4.5 Within Lab Variability Modified by Removal of High Values

	Original and Adjusted Variability, Within Lab							
			UDOT		PEPG		UofU	
Mix	Sample Cond.	Binder Grade	Original	Adjust.	Original	Adjust.	Original	Adjust.
A	62 mm Cut	64-34	14.1%		21.6%	10.3%	33.5%	17.6%
	62 mm Uncut		28.9%					
A	62 mm Cut	70-28	13.2%					
B	62 mm Cut	64-34	48.7%	27.2%	21.2%		25.2%	
	62 mm Uncut		28.9%		42.6%		24.0%	
	75 mm Uncut		13.5%		21.3%		19.7%	
B	62 mm Cut	70-28	18.4%		31.1%		25.9%	
	62 mm Uncut		25.1%		58.6%		31.4%	28.1%

4.4.2.2 Within-Lab and Between-Lab results

With values greater than two standard deviations removed, the within- and between-lab results are shown using the CoV to compare within-lab results and the ANOVA single factor test to evaluate whether the values created in different labs are derived from the same population (i.e., are the means and the variation consistent with one normal distribution or multiple distributions?). The results of these comparisons are given in Table 4.6.

Table 4.6 Variability Within and Between Labs

	Variability Within and Between Lab(s)					
Mix (Binder)	**Geometry**	UDOT	PEPG	UofU	Variability	ANOVA
A (64-34)	62 mm Cut	14.1%	10.3%	17.6%	Lower	Different
	62 mm Uncut	28.9%		33.5	Higher	NA
A (70-28)	62 mm Cut	13.2%				NA
B (64-34)	62 mm Cut	27.2%	21.2%	25.2%	Medium	Same
	62 mm Uncut	28.9%	42.6%	24.0%	Highest	Different
	75 mm Uncut	13.5%	21.3%	19.7%	Lowest	Same
B (70-28)	62 mm Cut	18.4%	31.1%	25.9%	Lower	Different
	62 mm Uncut	25.1%	58.6%	28.1%	Higher	Same

In the ANOVA column of the above table, if the between-lab data comes from the same population, the designation is ‘Same”. If not, the designation is “Different”. The following observations are available from this analysis:

- The adjusted coefficient of variation is below 25% except for the 62-mm uncut samples.
- Some inter-laboratory variability exists with the 62-mm cut samples independent of mix. This leads to questions about procedure.
- 62-mm uncut samples have the highest variability, independent of mix.
- 62-mm cut samples have acceptable variability.
- 75-mm uncut samples have the lowest variability; however, there were fewer replicates done with this configuration.

4.5 Comparison of results

The coefficient of variation from the I-FIT and the IDEAL-CT tests are listed in Table 4.7. These comparisons are done with data derived from 62-mm cut pucks and discarding values falling above two standard deviations from the mean. This procedure was used to normalize the distributions in both data sets.

Table 4.7 CoV Comparison I-FIT to IDEAL-CT 62mm Cut

Cracking Index Coefficent of Variation 62 mm Cut						
	Binder Grade	**Index Test**		**UDOT**	**PEPG**	**UofU**
Mix A	64-34	I-FIT	C of V	--	10%	14%
		IDEAL 62mm	C of V	14.1%	10.3%	17.6%
	70-28	I-FIT	C of V	--	--	13%
		IDEAL 62mm	C of V	13.2%	--	--
Mix B	64-34	I-FIT	C of V	25%	21%	27%
		IDEAL 62mm	C of V	27.2%	21.2%	25.2%
	70-28	I-FIT	C of V	26%	31%	18%
		IDEAL 62mm	C of V	18.4%	31.1%	25.9%
All specimens are cut to 62 mm height with target air voids at 7 ± 0.5%						

Several observations can be made:

- Coefficient of variation with both tests occasionally exceeds the target value of 25% in this sample configuration.
- The variability is approximately the same for both tests.
- No lab is always more (or less) variable than the others.

4.6 Observation and Discussion

The data indicates that to maintain the variability below a reasonable value, the procedure requiring a 62-mm puck compacted to 7±0.5% air voids to match the Hamburg Wheel Tracking (HWT) device specimen should not be used. Based on the limited data available in this study, testing samples compacted to 75 mm results in lower variability. There is some indication that compacting samples to 62 mm requires a large number of gyrations (higher than Ndes) to achieve height at the target density. Since all UDOT production mixes achieve 3.5% air void at Ndes, it seems unlikely that the contracting community would build this harsh of a mix in practice. Both of the lab mixes were adapted from mixes containing 20% RAP using a 64-34 binder and are not verified mix designs in their present forms. It is unknown at present what the comparative variability between 62-mm and 75-mm uncut samples of production mixes would be.

When pucks are cut to height in either test, the cross section is more consistent. Less variation is present due to the comparative surface smoothness vs a raw compacted specimen. The roughness of the surface has a smaller relative impact on the cross section of a 75-mm puck vs a 62-mm puck. All cut specimens are derived from pucks which are at least 115 mm tall. If the particles are having difficulty orienting in a thinner specimen, the compactive effort is fracturing aggregates in a random manner creating unbound surfaces and increasing variability. It is a good idea to build both the HWT specimens and the Cracking Index specimens in the same format so that they can be randomly selected as to test.

Based on the data, both the I-FIT and the IDEAL tests result in approximately the same variability. This variability can, and should, be held below a CoV of 25%. In a normal distribution 95.5% of all data falls within two standard deviations of the mean, and 68.3% falls within one standard deviation. This statistical information should be used to develop a process by which the validity of the data could be assessed based on 3 or 4 samples. Once enough data for these tests has been collected, a procedure can be developed in which the two closest results are averaged and then the third value is rejected if it is further away from the average than the standard deviations multiplied by a value. If this rejection is necessary, one additional sample should be prepared, tested, and the results used to calculate a new average. This places three measured values inside a standard distribution curve with known variation to ensure a CoV below an acceptable value. As more data becomes available, it will be possible to state if the need to reject a sample is a rare occurrence or not. This approach is analogous to the procedures specified in ASTM C39 for testing concrete specimens.

Chapter 5 Study of aging conditions from plant to laydown to field cores

5.1 Overview

In this study, data from different mixtures at three different conditions, starting at the plant, then at laydown, and from cores obtained after three years of service was obtained. Such unique data allows for an analysis of how the mechanical properties of the material are affected by aging conditions. Looking at the progression of material properties with time will allow setting of a threshold, or limit, that can be used to prevent premature failure. Other relevant information includes the relation between material variability and expected performance as well as the relation between different tests.

5.2 Comparison of data

For each mixture, data were obtained from laboratory-prepared samples and from field cores. It is known that differences in compaction as well as differences in air voids could affect the results and increase the variability. Furthermore, after being exposed to the environment for several years, the chemical composition of the material has changed due to oxidative aging. Therefore, comparisons between lab compacted and core data must be done with caution. The effect of aging is further discussed in this study.

5.2.1   Flexibility index

The Flexibility Index, FI, obtained from the different mixtures at the different aging conditions is shown in Figure 5.1. The figure is based on the complete set of data from the cores (i.e., highest values not removed). The figure shows that sections UT-02 and UT-05 both have the lowest flexibility index values; these two sections are expected to show poor cracking performance based on FI values lower than 8. In all sections, the effect of aging is obvious since the FI value from 3-year old cores is lower than it was during laydown.

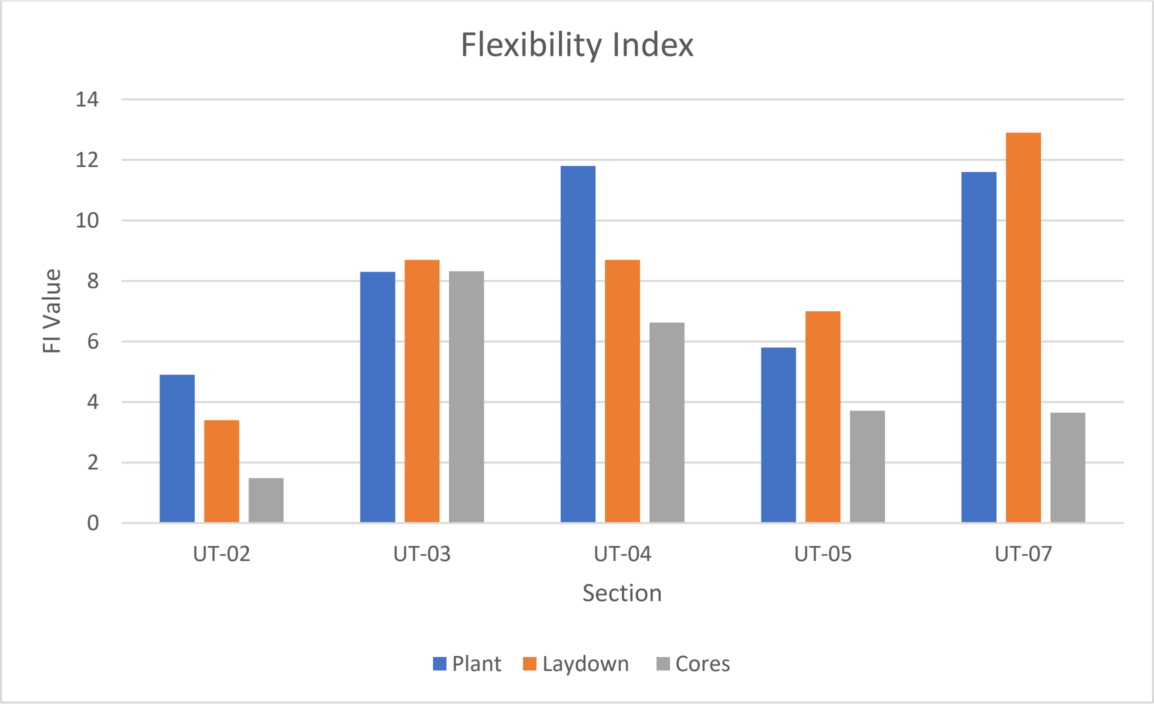


Figure 5.1 Comparison of FI Data at 25 ºC

Figure 5.1 shows that there is a general trend of decreasing FI from the plant to laydown (short-term aging) and from laydown to cores (long-term aging). Regardless of the aging conditions, Section UT-02 is expected to have the worst performance of the group and using a threshold of 8 would also place UT-05 as a potential low performer. These two mixtures were the only ones of the group designed following Marshall procedures and thus are not ‘UDOT-type’ mixtures.

Section UT-03 is the most consistent (i.e., no aging effects) with negligible changes in FI values across different aging periods.. Section UT-07 had the highest FI during mixing and laydown, but had one of the lowest FI values in cores. It is hypothesized that the reason for such significant effect in aging might be related to its high virgin binder content. The decrease in FI should be reflected in future performance.

It is also noted that those sections with low flexibility index also have high variability. UT-02, UT-05, and UT-07 had coefficient of variation greater than 20% even after correcting for possible outliers.

5.2.2 Comparison between FI and CT index

The results from the comparison of the CT Index and the iFIT for cores are shown in Figure 5.2. In the same way as the other tests, the CT index predicts that section UT-02 will have the worst performance of the group while section UT-03 is predicted to have the best performance of the group. However, unlike the two other tests, the CT Index shows section UT-04 as likely a poor performer and UT-05 should have better performance. This is the only test that predicts poor performance for section UT-04; the two other tests show a significant decrease in performance between laydown and coring.

The variability in the results is also very high for sections UT-02, UT-03, and UT-07 having coefficient of variation greater than 40%, while for sections UT-04 and UT-05 the variability is 11% and 6%, respectively.

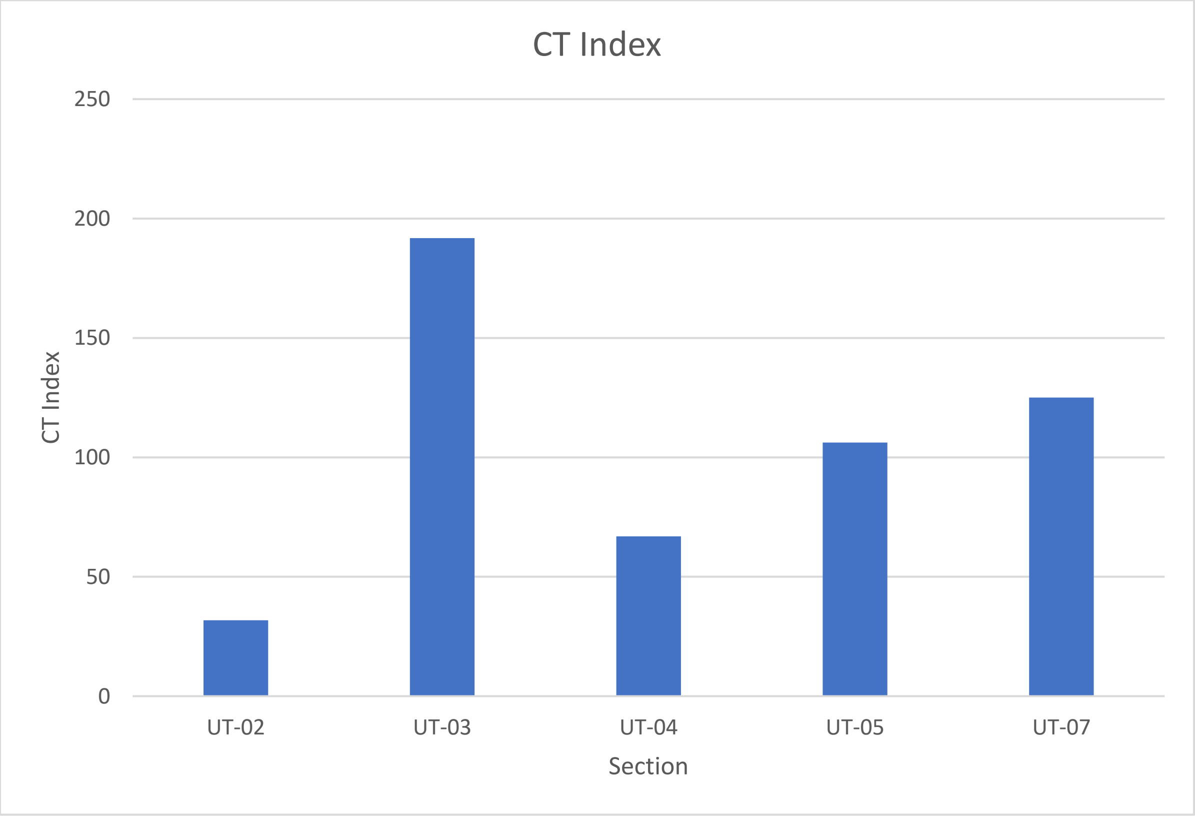


Figure 5.2 CT Index at 25 ºC

Given that CT Index is a newer test, there is no previous data to use as a reference, and to take advantage of the wealth of knowledge previously accumulated using the IFIT test, a relative comparison between the FI and the CT-Index was made. It is understood that the numbers would be different, but since both tests are based on similar concepts, it would be expected that a strong trend exists between them. The comparison between FI and CT Index is shown in Figure 5.3.

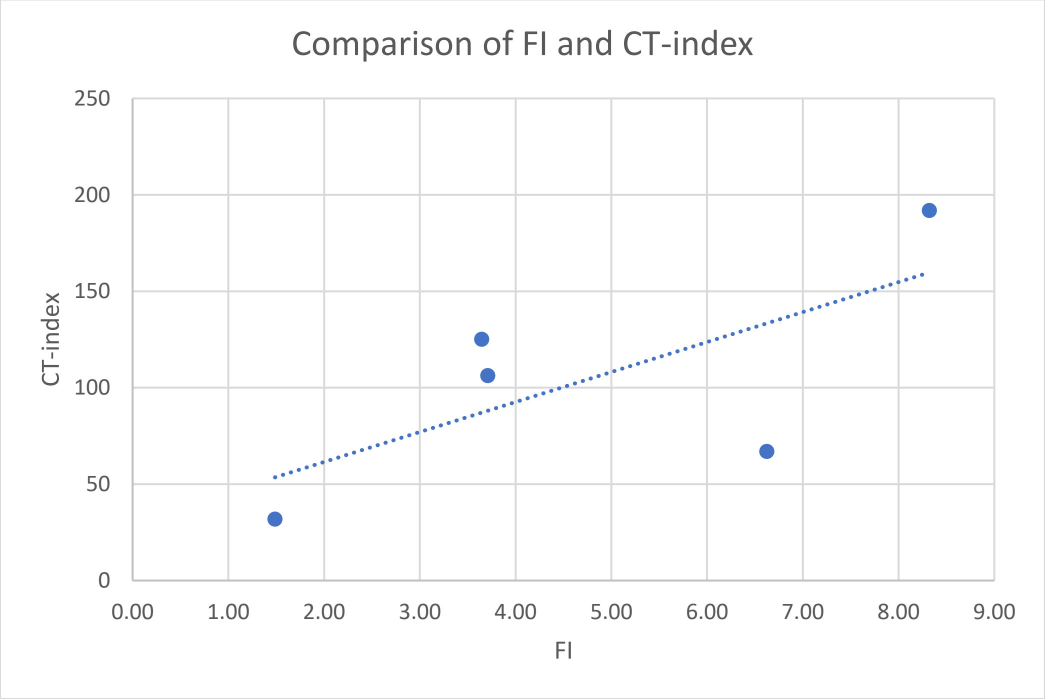


Figure 5.3 Comparison Between FI and CT-Index

Figure 5.3 shows that, for cores, the FI and CT Index have the same prediction for the best and worst expected performance; however, there is no agreement with intermediate performance predictions.

An alternative way to compare results between tests is shown in Table 5.1 in which the predicted best performing sections are shown. Similarly, Table 5.2 shows the predicted worst performers.

Table 5.1 Predicted Best Performers

2017 Lab compacted		**2020 Cores**	
FI		**FI**	**CT Index**
Plant	Laydown		
UT-04	UT-07	UT-03	UT-03
UT-07	UT-03/04	UT-04	UT-07

Table 5.2 Predicted Worst Performers

2017 Lab compacted		**2020 Cores**	
FI		**FI**	**CT Index**
Plant	Laydown		
UT-02	UT-02	UT-02	UT-02
UT-05	UT-05	UT-07	UT-04

Tables 5.1 and 5.2 show that there are differences among the tests in the prediction of the sections that are expected to perform well and the ones that do not. The only real commonality is that both FI and CT

indices predict that section UT-02 should have poor performance. This holds true regardless of the aging condition evaluated.

5.3 Summary

Different tests were performed on asphalt mixtures obtained during mixing and compaction and from cores after 3 years on the road. All tests predicted that the same section, UT-02, would have poor performance.

All tests also predict that another section, UT-03, would have the best performance. There was, however, no agreement amongst the different tests for the ‘intermediate’ performing sections. It is not clear if index-type tests, such as the ones performed as part of this research, are meant for such fine-tuned predictions instead of providing pass-fail information.

Chapter 6. Conclusions and Recommendations

The goal and objective of this project is to develop an implementable holistic methodology for improving asphalt mix durability in all three aspects: lab mix design, plant production QC/QA, and field placement. A comprehensive literature review was performed regarding the efforts in ensuring the durability and performance of asphalt mixture in mix design and production. Based on the solution and challenges mentioned in literature review, a coherent framework of holistic methodology for improving asphalt mix durability in the areas of lab mix design, plant production, and field placement is proposed. Subsequent research and performance testing will be carried out around the framework. According to the challenges summarized in the review, relative studies are conducted to solve them before starting test parts, including variability of performance tests and effect of aging conditions on mechanical properties of the material. This summarizes of findings from the Year 1 of the research efforts are listed as follow:

- The development of the performance specification and performance mix design is summarized. Four approaches in balanced mix design/index-based performance mix design and one predictive performance mix design method are demonstrated. The five approaches are compared based on six evaluation criteria proposed in this review.
- The challenges and solutions in incorporating performance tests in asphalt productions are discussed, including selecting testing methods, determining the appropriate index threshold limits, sampling position, and testing frequency et al.
- A coherent BMD/QC/QA framework was established that has four interconnected components: (a) volumetric mix design and selection of multiple asphalt contents for mixture performance evaluation, (b) mixture performance evaluation at multiple asphalt contents and selection of the balanced asphalt content, (c) mixture performance verification at the balanced asphalt content, and (d) production QC/QA testing.
- One set of laboratory tests and associated acceptance criteria are recommended for the mixture performance evaluation and the production QC/QA testing and the other set is recommended for the mixture performance verification wherein DOTs can choose their preferred performance verification tests.
- Two practical loose-mixture aging protocols were developed, one for short-term aging used in the process of the volumetric mix design, the mixture performance evaluation, and the production QC/QA testing, and the other for mid-term aging employed in the mixture performance verification. The

recommended short-term aging protocol is to age the loose mixture in a force draft oven for 2 hr at the mixture compaction temperature, while the mid-term aging protocol consists of three steps: (1) short-term aging, (2) 20-hr loose-mixture aging at 100°C, and (3) reheating for compaction. A case study was presented in this report to demonstrate the whole process of the framework, including the actual plant production QC testing.

- The variability of main performance tests in mix design were investigated including SCB, I-FIT, and IDEAL-CT. Based on experience with asphalt mixtures testing, a coefficient of variation less than 25% is desirable to assure that the results are meaningful for acceptance testing. When pucks are cut to height in either test, the cross section is more consistent. Less variation is present due to the comparative surface smoothness vs a raw compacted specimen.
- Based on the data, both the I-FIT and the IDEAL tests result in approximately the same variability. In a normal distribution 95.5% of all data falls within two standard deviations of the mean, and 68.3% falls within one standard deviation. This statistical information should be used to develop a process by which the validity of the data could be assessed based on 3 or 4 samples. Once enough data for these tests has been collected, a procedure can be developed in which the two closest results are averaged and then the third value is rejected if it is further away from the average than the standard deviations times a value.
- To determine how the mechanical properties of the material are affected by aging conditions, data from different mixtures at three different conditions, starting at the plant, then at laydown, and from cores were obtained after three years of service. Looking at the progression of material properties with time will allow setting of a threshold, or limit that can be used to prevent premature failure.
- Three different tests were performed on asphalt mixtures obtained during mixing and compaction and from cores after 3 years on the road. All tests predicted that section UT-02 would have poor performance. The IFIT predicted that section UT-05 should have poor performance; one test based the prediction on the low m-value and the other one based it on the low FI. Looking at the data obtained from the cores, the FI predicts that after three more years, section UT-07 will start to deteriorate. All three tests also predict that section UT-03 would have the best performance. There was, however, no agreement amongst the different tests for the ‘intermediate’ performing sections. It is not clear if index-type tests, such as the ones performed as part of this research, are meant for such fine-tuned predictions instead of providing pass-fail information.

References

AASHTO (2003). Major Types of Transportation Construction Specifications: A Guideline to Understand Their Evolution and Application. Report of the AASHTO Highway Subcommittee on Construction. American

Association of State Highway and Transportation Officials. Washington, D.C.

Abdalla, A., Faheem, A., Hosseini, A., and Titi, H. (2021). Performance Related Asphalt Mixtures Characterization. Conference Presentation, Transportation Research Board 100th Annual Meeting. January 2021. Washington DC, United States.

Advanced Asphalt Technologies (AAT), LLC, (2011). NCHRP Report 673: Manual for Design of Hot Mix Asphalt with Commentary. Transportation Research Board, Washington, D.C.

Alabama Department of Transportation (ALDOT) (2020). Balanced Asphalt Mix Design for Local Roads. Special Provisions 18-0806(2). March 21, 2020.

Al-Khayat, H., and Epps Martin, A. (2021). Balanced Mix Design Approach for Superpave Hot-Mix Asphalt (HMA) Mixtures with RAP. Conference Presentation, Transportation Research Board 100th Annual Meeting. January 2021. Washington DC, United States.

Al-Qadi, I. L., Said, I. M., Ali, U. M., and Kaddo, J. R. (2021). The Truth and Myth about the Cracking Mechanism of Asphalt Concrete. Conference Presentation, Transportation Research Board 100th Annual Meeting. January 2021. Washington DC, United States.

ARA (2004). Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures. Final Report of NCHPR 1-37A Project. Transportation Research Board of the National Academies, Washington, D.C.

Aschenbrener, T., and N. Far (1994), Short-Term Aging of Hot Mix Asphalt, Colorado Department of Transportation Public Report No. CDOT-DTD-R-94-11.

Ashouri, M., Wang, Y. D., Choi, Y., and Kim, Y. R. (2021). Development of healing model and simplified characterization test procedure for asphalt concrete. Construction and Building Materials, Vol. 271. DOI: 10.1016/j.conbuildmat.2020.121515.

Asphalt Institute (2015). MS-2 Asphalt Mix Design Methods (7th edition).

Bell, C. A., Y. AbWahab, M. Cristi, and D. Sosnovske (1994), Laboratory Aging of Asphalt-Aggregate Mixtures: Field Validation, Strategic Highway Research Program, SHRP-A-390, National Research Council, Washington, D.C.

Bennert, T. (2011), Implementation of Performance-Based HMA Specialty Mixtures in New Jersey, Journal of the Association of Asphalt Paving Technologists, Vol. 80, pp. 719–740.

Bennert, T. (2020). New York's State's Balanced Mixture Design (BMD) Research Efforts. Conference Presentation. Northeast Asphalt Users Producer Group (NEAUPG). October 28, 2020.

Bennert, T., E. Hass, E. Wass Jr., and B. Berger (2020), Indirect Tensile Testing for Balanced Mixture Design and Quality Control Performance Testing, Journal of Association of Asphalt Paving Technologists.

Bennert, T., Sheehy, E., Blight, R., Gresavage, S., and Fee, F. (2014). Implementation of Performance-Based Specifications for Asphalt Mix Design and Production Quality Control for New Jersey**.** Circular Number E-C189, Transportation Research Board of the National Academies, Washington, D.C.

Braham, A. F., W. Buttlar, T. Clyne, M. Marasteanu, and M. Turos (2009), The Effect of Long-term Laboratory Aging on Asphalt Concrete Fracture Energy, Journal of the Association of Asphalt Paving Technologists (AAPT), Vol. 78.

Buchanan, S. (2016). Balanced Mix Design (BMD) for Asphalt Mixtures. Conference Presentation. http://shrp2.transportation.org/documents/presentations/10_Buchanan_SHRP2BalancedMixDesignforAsphaltMixtures.pdf. Accessed in August 2021.

Buttlar, W. G., B. Hill, H. Wang, and W. Mogawer (2016), Performance space diagram for the evaluation of high- and low-temperature asphalt mixture performance, Journal of the Association of Asphalt Paving Technologists, 2016, Vol. 85.

Buttlar, W. G., Jahangiri, B., Rath, P., Majidifard, H., Urrar, L., Meister, J., and Brown, H. (2021). Development of Performance-Related Asphalt Mix Design Specification for the Illinois Tollway. Final Report. Illinois State Toll Highway Authority.

Buttlar, W. G., Urra-Contreras, L., Jahangiri, B., Rath, P., and Majidifard, H. (2020). Support for Balanced Asphalt Mixture Design Specification Development in Missouri. Final Report Missouri Department of Transportation. CMR 20-010.

Chen, C., F. Yin, P. Turner, R. C. West, and N. Tran (2018), Selecting a Laboratory Loose Mix Aging Protocol for the NCAT Top-Down Cracking Experiment, Transportation Research Record: Journal of the Transportation Research Board, Vol. 2772(28), pp. 359–371.

Christensen, D., and R. Bonaquist (2002), Use of Strength Tests for Evaluating the Rut Resistance of Asphalt Concrete, Journal of the Association of Asphalt Paving Technologists, Vol. 71, pp. 692–711.

Christensen, Jr., D. W. and Bonaquist, R. F. (2006). NCHRP Report 567: Volumtric Requirements for Superpave Mix Design. Transportation Research Board, Washington, D.C.

Coleri, E., Sreedhar, S., and Obaid, I. A. (2020). Development of a Balanced Mix Design Method in Oregon. Final Report. Oregon Department of Transportation. FHWA-OR-RD-21-03.

Cross, S. and Li, J. (2019). Implement Balanced Asphalt Mix Design in Oklahoma. Final Report.

Oklahoma Department of Transportation. FHWA-OK-19-01.

Daniel, J. S., Corrigan, M., Jacques, C., Nemati, R., Dave, E., and Congalton A. (2018). Comparison of Asphalt Mixture Specimen Fabrication Methods and Binder Tests for Cracking Evaluation of Field Mixtures. Road Materials and Pavement Design. DOI: 10.1080/14680629.2018.1431148

Dave, E. V., M. Oshone, A. Schokker, and C. Bennett (2019), Disc Shaped Compact Tension (DCT) Specifications Development for Asphalt Pavement, MN/RC 2019-24, Minnesota Department of Transportation, St. Paul, Minnesota.

Diffenderfer, S. D. and Bowers, B. (2019). Initial Approach to Performance (Balanced) Mix Design: The Virginia Experience. Transportation Research Record. Vol. 2673 (2), pp: 335-345.

Diffenderfer, S. D., Boz, I., Habbouche, J., (2021). Balanced Mix Design for Surface Asphalt Mixtures: Phase I: Initial Roadmap Development and Specification Verification. Final Report. Virginia Department of Transportation. FHWA/VTRC 21-R15.

Diffenderfer, S., Boz, I., Habbouche, B. (2020). Round Robin Phase 1: Evaluating Indirect Tensile Test Data (IDT) for CT-index. https://vaa.mynewscenter.org/round-robin-phase-1-evaluating-indirect-tensile-test-data-idt-for-ct-index/. Accessed in August 2021.

Ding, J., Wang, Y. D., Gulzar, S., Kim, Y. R., and Underwood, B., S. (2020). Uncertainty Quantification of Simplified Viscoelastic Continuum Damage Fatigue Mode Using the Bayesian Inference-Based Markov Chain Mont-Carlo Method. Transportation Research Record. 2674(4). DOI: 10.1177/0361198120910149

Epps, J. A., Hand, A., Seeds, S., Schulz, T., Alavi, S., Ashmore, C., Monismith, C. L., Deacon, J. A., Harvey, J. T., and Leahy, R. (2002). NCHRP Report 455: Recommended Performance-Related Specification for Hot-Mix Asphalt Construction: Results of the WesTrack Project. Transportation Research Board of the National Academies, Washington, D.C.

Epps-Martin, A., E. Arambula, F. Yin, L. Garcia Cucalon, et al. (2014), NCHRP Report 763: Evaluation of the Moisture Susceptibility of WMA Technologies, Transportation Research Board of the National Academies, Washington, D.C.

Etheridge, R. A., Wang, Y. D., Kim, S. S., and Kim, Y. R. (2019). Evaluation of Fatigue Cracking Resistance of Asphalt Mixtures Using Apparent Damage Capacity. Journal of Materials in Civil Engineering. 31(11).

Federal Highway Administration (FHWA) (2019). Performance Engineered Pavements. Tech Brief. Federal Highway Administration. FHWA-HIF-20-005.

Fugro Consultants, Inc. and Arizona State University. (2011). A Performance-Related Specification for Hot-Mixed Asphalt, NCHRP Report 704. Transportation Research Board, Washington, DC.

Garcia-Ruiz, J., and Sargand, S. (2021). Rational Approach to Evaluate Asphalt Concrete Base Course

Design for Improving Construction Quality and Performance. Conference Presentation, Transportation Research Board 100th Annual Meeting. January 2021. Washington DC, United States.

Golalipour, A., Veginati, V., and Mensching, D. (2021). Evaluation of Asphalt Mixture Performance Using Cracking and Durability Tests at a Full-Scale Pavement Facility. Conference Presentation, Transportation Research Board 100th Annual Meeting. January 2021. Washington DC, United States.

Hajj, E. Y., A. Hand, R. Chkaiban, and T. Aschenbrener (2019), Index-Based Tests for Performance Engineered Mixture Designs for Asphalt Pavements, FHWA-HIF-19-103, FHWA.

Harvey, J., Wu, R., Signore, J., Basheer, I., Holikatti, S., Vacura., P., and Holland, T. J. (2014). Performance-Based Specificatioin: California Experience to Date. Circular Number E-C189, Transportation Research Board of the National Academies, Washington, D.C.

Hicks, R. G. (1991), Moisture Damage in Asphalt Concrete, NCHRP Synthesis of Highway Practice 175, Transportation Research Board, Washington, D.C.

Ho, C.H. and Romero, P: (2012) Asphalt Mixture Beams Used in Bending Beam Rheometer for Quality Control Paper 12-1273 Journal of The Transportation Research Board, No 2268, Washington, D.C. Pp. 92-97.

Hu, S., F. Zhou, and T. Scullion (2014), Development of Texas Mechanistic-Empirical Flexible Pavement Design System (TxME), FHWA/TX-14/0-6622-2, Texas A&M Transportation Institute, College Station, Texas.

Hughes, C. (2005), NCHRP Synthesis 346: State Construction Quality Assurance Programs. Transportation Research Board of the National Academies, Washington, D.C.

Hughes, C. S., Moulthrop, J. S., Tayabji, S., Weed, R. M., and Burati, J. L. (2012). NCHRP Research Results Digest 371: Quality-Related Pay Adjustment Factors for Pavements. Transportation Research Board of the National Academies Washington, D.C.

Illinois Department of Transportation (IDOT) (2021). Special Provision for Hot-Mix-Asphalt – Mixture Design Verification and Production (Modified for I-FIT).

Im, S., P. Karki, and F. Zhou (2016). Development of a New Mix Design Method for Asphalt Mixtures Containing RAP and Rejuvenators, Construction and Building Materials, Vol. 115, pp. 727-734.

Im, S., S. Hu, and F. Zhou (2015), Performance Studies and Future Directions for Mixes Containing RAP/RAS: Technical Report, FHWA/TX-15/0-6738-2, Texas A&M Transportation Institute, College Station, Texas.

Jeong, J., Y. D. Wang, A. Ghabari, C. Nash, D. Nener-Plante, B. S. Underwood, and Y. R. Kim. (2020). Pavement Performance Predictions Using Performance-Volumetric Relationship and Evaluation of Construction Variability: Example of MaineDOT Shadow Project for the Development of Performance-Related Specifications. Construction and Building Materials. Vol. 263. DOI:

https://doi.org/10.1016/j.conbuildmat.2020.120150.
Johanneck, L, Geib, J., Van Deusen, D., Garrity, J., Hanson, C., and Dave, E. (2015). DCT Low Temperature Fracture Testing Pilot Project. Minnesota Department of Transportation. MN/RC 2015-20.

Karki, P., and F. Zhou (2018), Evaluation of Asphalt Binder Performance with Laboratory and Field Test Sections, FHWA/TX-18/0-6674-01-R1, Texas A&M Transportation Institute, College Station, Texas.

Kim, Y. R., Castorena, C., Elwardany, M., Fousefi Rad, F., Underwood, S., Gundla, A., Gudipudi, P., Farrar, M., and Glaser, R. (2018). Long-Term Aging of Asphalt Mixtures for Performance Testing and Prediction. NCHRP Report 871. Transportation Research Board, Washington, DC.

Kim, Y. R., Guddati, M., Choi, Y., Kim, D., Norouzi, A., Wang, Y. D., Keshavarzi, B., Ashouri, M., Ghabari, A., Wargo, A. D. (2022). Hot-Mix Asphalt Performance-Related Specification Based on Viscoelastoplastic Continuum Damage (VEPCD) Models. Final Report, FHWA-HRT-21-093. Federal Highway Administration, McLean, VA.

Kim, Y. R., Underwood, B. S., Guddati, M. N., Ghanbari, A., Wang, Y. D., Keshavarzi, B., Jeong, J., Mocelin, D., Pivetta, F., and Saleh, N. (2022). Development and Deploy Performance Related Specifications (PRS) for Asphalt Pavement Construction. Final Report. Federal Highway Administration, McLean, VA. In Press.

Ksaibati, K., and Butts, N. (2003). Evaluating the Impact of QC/QA Programs on Asphalt Mixture Variability. MPC Report No. 03-146, Mountain-Plains Consortium, Fargo, ND.

Lee, J., Haddock, J. E., Alvarez, D., Rastegar, R. R. (2020). Quality Control and Quality Assruance of Asphalt Mixtgure Using Laboratory Rutting and Cracking Test. Final Report, Indiana Department of Transportation and Purdue University. FHWA/IN/JTRP-2019/19.

Liu, J., Zhao, S., and Saboundjian, S., (2017). Variability of Composition, Volumetric, and Mechanic Properties of Hot Mix Asphalt for Quality Assurance. Journal of Materials in Civil Engineering, 29(2).

Louisiana Department of Transportation and Development (LDOTD) (2016). Louisiana Standard Specifications for Roads and Bridges 2016 Edition.

Maupin, G. W. (2003). Additional Asphalt to Increase The Durability of Virginia's Superpave Surface Mixes. Final Report. Virginia Transportation Research Council. VTRC 03-R15.

McCarthy, L. M., Callans, J., Quigley, R., and Scott, III, S. V. (2016). NCHRP Synthesis 492: Performance Specifications for Asphalt Mixtures: A Synthesis of Highway Practice. Transportation Research Board of the National Academies, Washington, D.C.

McDaniel, R. S., Leahy, R. B., Huber, G., Moulthrop, J. S., and Ferragut, T.. (2011). NCHRP Web-Only Document 186: The Superpave Mix Design System: Anatomy of a Research Program. Transportation Research Board, Washington, D.C. DOI: http://dx.doi.org/ 10.17226/22812.

Missouri Department of Transportation (MoDOT) (2021a). Balanced Mix Design Performance Testing and Increased Density. Job Special Provisions. NJSP-20-01B.

Missouri Department of Transportation (MoDOT) (2021b). Balanced Mix Design Performance Testing for Job Mix Approval. Job Special Provision. NJSP-21-08.

Mogawer, W. S., A. Austerman, K. Stuart, F. Zhou, and P. Romero (2019), Balanced Mix Design Sensitivity to Production Tolerance Limits and Binder Source, Journal of the Association of Asphalt Paving Technologists, Volume 88, pp. 857–876.

Mogawer, W. S., Austerman, A. J., Kluttz, R., and Roussel, M. (2012). High-Performance Thin-Lift Overlays with High Reclaimed Asphalt Pavement Content and Warm-Mix Asphalt Technology: Performance and Workability Characteristics. Transportation Research Record, Vol 2293.

Mohammad, L. N., and S. Cooper III (2016), Implementation of a Balanced Asphalt mixture Design Procedure: Louisiana's Approach, Journal of the Association of Asphalt Paving Technologists, Volume 85, 2016, pp. 857–876.

Mohammad, L. N., Elseifi, M. A., Cooper, S. B., Hughes, C. S., Button, J. W., and Ervin L. D., Jr. (2016a). Comparing the Volumetric and Mechanical Properties of Laboratory and Field Specimens of Asphalt Concrete. NCHRPT Report 818. Transportation Research Board, Washington, D.C.

Mohammad, L., Kim, M., and Challa, H. (2016b). Development of Performance-based Specification for Louisiana Asphalt Mixtures. Final Report, FHWA/LA.14/558. Louisiana Transportation Research Center, Baton Rouge, Louisiana.

Montoya, M. A., Pouranian, M. R., Haddock, J. E. (2018). Increasing Asphalt Pavement Density Through Mixture Design: A Field Project. Journal of Association of Asphalt Paving Technologists.

National Asphalt Pavement Association (NAPA) (2019). Market Facts. Accessed in August 2019. http://www.asphaltpavement.org/index.php?option=com_content&view=article&id=891
National Cooperative Highway Research Program (NCHRP) (2004). NCHRP Research Results Digest 291: Quality Characteristics for Use with Performance-Related Specifications for Hot Mix Asphalt, Transportation Research Board of the National Academies, Washington, D.C.

New Jersey Department of Transportation (NJDOT), 2011. NJDOT Survey on Dust to Binder Ratio for Superpave.http://www.pavementinteractive.org/wp-content/uploads/2011/01/SuperpaveDBRatioSurvey.pdf. Accessed in June, 2019.

Newcomb, D. and F. Zhou (2018), Balanced Design of Asphalt Mixtures, MN/RC 2018-22, Minnesota Department of Transportation, St. Paul, Minnesota.

Newcomb, D. E., Arambula, E., Yin, F., Zhang, J., Bhasin, A., Li, W., and Arega, Z. (2015). Properties of Foamed Asphalt for Warm Mix Asphalt Applications. NCHRP Report 807. Transportation Research Board of the National Academies, Washington, D.C.

Newcomb, D., A. Epps-Martin, F. Yin, E. Arambula, et al. (2015), Short-Term Laboratory Conditioning of Asphalt Mixtures, NCHRP Report 815, Transportation Research Board, Washington, D.C.

Nsengiyumva, G., Kim, Y, and Hu, J. (2020). Feasibility and Implementation of Balanced Mix Design in Nebraska. Final Report. Texas A&M University. SPR-PI(19) M080.

Ozer, H. and I. Al-Qadi (2018), Development and Implementation of the Illinois Flexibility Index Test A Protocol to Evaluate the Cracking Resistance of Asphalt Mixtures, Transportation Research Circular E-C237: Innovations in Asphalt Mixture Design Procedures, TRB.

Paye, B., "WisDOT Pilot High Recycle Mixes," Wisconsin Asphalt Pavement Association (WAPA) Annual Conference, Dec. 2014 [Online]. Available: http://www.wispave.org/wp-content/uploads/dlm_uploads/Breakout-2014-Paye-High-Recycle.pdf [accessed Sept. 1, 2015].

Petersen, J. C. (2009), Transportation Research E-Circular (E-C140): A Review of the Fundamentals of Asphalt Oxidation: Chemical, Physicochemical, Physical Property, and Durability Relationship, Transportation Research Board of the National Academies, Washington, D.C.

Prowell, B., and E. Brown (2007), Superpave Mix Design: Verifying Gyration Levels in the Ndesign Table, NCHRP Report 573, Transportation Research Board of the National Academies, Washington, D.C.

Reinke, G., A. Hanz, and D. Herlitzka, (2015), Further Investigations into the Impact of REOB & Paraffinic Oils on the Performance of Bituminous Mixtures, presentation at FHWA ETG Meeting, Fall River, MA, April 2015.

Romero, P., and Van Frank, K. (2017). Balanced Asphalt Concrete Mix Performance Phase II: Analysis of BBR and SCB-IFIT Tests. Report UT17.21 Utah Department of Transportation, Salt Lake City, UT

Romero, P., and Van Frank, K. (2019). Balanced Asphalt Concrete Mix Performance Phase III: Evaluation of Field Materials Using the BBR and SCB-IFIT Tests. Report UT19.15 Utah Department of Transportation, Salt Lake City, UT

Romero, P. (2021). Balanced Asphalt Concrete Mix Performance Phase V: Field Evaluation for Intermediate and Low Temperature Cracking. Report UT21.24. Utah Department of Transportation, Salt Lake City, UT.

Scott S., Konrath, L., Ferragut, T., and Loulakis, M. (2013). Performance Specifications for Rapid Highway Renewal, Report No. S2-R07-RR-1, Transportation Research Board, Washington, DC.

Scott, S. (2019). Improving Durability of Asphalt Mixtures. Final Report. Colorado Department of Transportation – Research. CDOT-2019-06.

Texas Department of Transportation (TxDOT) (2019). Special Specification 3074: Superpave Mixtures – Balanced Mix Design.

Tran, N., G. Huber, F. Leiva, et al. (2019), Mix Design Strategies for Improving Asphalt Mixture Performance, NCAT Report 19-08, National Center for Asphalt Technology, Auburn, Al.

Tran, N., Yin, F., Leiva, F., Rodezno, C., Huber, G., and Pine, B., Adjustments to the Superpave Volumetric Mixture Design Procedure for Selecting Optimum Asphalt Content. Final Report of NCHRP 20-07/Task 412. Transportation Research Board, Washington, DC.

Transportation Research Board (TRB) (2005). Glossary of Highway Quality Assurance Terms. Transportation Research Circular, Number E-C074. Transportation Research Board, Washington, D.C.

Transportation Research Board (TRB) (2018). Glossary of Transportation Construction Quality Assurance Terms, Seventh Edition. Transportation Research Circular, Number E-C235. Transportation Research Board, Washington, D.C.

Tsai, B. W., R. Wu, J. Harvey, and C. Monismith. Development of Fatigue Performance Specification and Its Relation to Mechanistic–Empirical Pavement Design Using Four-Point Bending Beam Test Results. Presented at 4-Point Bending, CRC/Balkema, Leiden, Netherlands, 2012.

TxDOT Construction Specifications (2014).

Ullidtz, P., J. Harvey, I. Basheer, D. Jones, R. Z. Wu, J. Lea, Q. Lu (2010), CalME: A New Mechanistic-Empirical Design Program for Flexible Pavement Rehabilitation, Journal of Transportation Research Record 2153, pp 143–152.

Underwood, B. S., Baek, C., and Kim, Y. R. (2012). Simplified Viscoelastic Continuum Damage Model as Platform for Asphalt Concrete Fatigue Analysis. Transportation Research Record, Vol. 2296, pp. 35-45.

Van Frank, K. and Romero, P. (2020). Balanced Asphalt Concrete Mix Performance Phase IV: Cracking Indices for Asphalt Mixtures. Report UT20.13 Utah Department of Transportation, Salt Lake City, UT
Vermont Agency of Transportation (VTrans) (2019). Superpave Bituminous Concrete Pavement, Performance Engineered Method.

Virginia Asphalt Association (VAA) (2021). BMD Round Robin: Phase 2 Begins. https://vaa.mynewscenter.org/bmd-round-robin-phase-2-begins/. Accessed in August 2021.

Wang, Y. D. (2019). Development of the Framework of Performance-Engineered Mixture Design for Asphalt Concrete. Ph.D. Dissertation. Department of Civil, Environmental, and Construction Engineering, North Carolina State University. Raleigh, North Carolina.

Wang, Y. D. and Kim, Y. R. (2017). Development of a Pseudo Strain Energy-Based Fatigue Failure Criterion for Asphalt Mixtures. International Journal of Pavement Engineering. DOI: 10.1080/10298436.2017.1394100

Wang, Y. D., B. Keshavarzi, and Y. R. Kim (2018), Fatigue performance prediction of asphalt pavements with FlexPAVETM, the S-VECD model, and DR failure criterion, Journal of Transportation Research Record, 2672 (40), 217–227.

Wang, Y. D., Ghanbari, A., Underwood, B. S., and Kim, R. Y. (2021a). Development of Framework of the Predictive Performance-Engineered Mix Design Procedure for Asphalt Mixtures. International Journal of Pavement Engineering. DOI: 10.1080/10298436.2021.1938044

Wang, Y. D., Ghanbari. A., Underwood, B. S., and Kim, Y. R. (2019). Development of a Performance-Volumetric Relationship for Asphalt Mixtures. Transportation Research Record. DOI: 10.1177/0361198119845364

Wang, Y. D., Ghanbari. A., Underwood, B. S., and Kim, Y. R. (2021b). Development of Preliminary Transfer Functions for FlexPAVE™. Construction and Building Materials, Vol. 266, Part B. DOI: 10.1016/j.conbuildmat.2020.121182

Wang, Y. D., Keshavarzi, B., and Kim, Y. R. (2018). Fatigue Performance Predictions of Asphalt Pavements Using FlexPAVE™ with the S-VECD Model and D^R Failure Criterion. Transportation Research Record, Vol. 2642(40). DOI: 10.1177/0361198118756873

Wang, Y. D., Norouzi, A., and Kim, Y. R. (2016). Comparison of Fatigue Cracking Performance of Asphalt Pavements Predicted by Pavement ME and LVECD Programs. Transportation Research Record, Vol. 2590 (1), 44-55. DOI: 10.3141/2590-06

Wang, Y. D., Underwood, B. S., and Kim, Y. R. (2020). Development of a Fatigue Index Parameter, Sapp, for Asphalt Mixes Using Viscoelastic Continuum Damage Theory. International Journal of Pavement Engineering. DOI: 10.1080/10298436.2020.1751844

Wen, H. F., S. H. Wu, and L. N. Mohammad. (2016), Long-Term Field Rutting and Moisture Susceptibility Performance of Warm Mix Asphalt Pavement, Transportation Research Record: Journal of the Transportation Research Board, No. 2575, Transportation Research Board of the National Academies, Washington, D.C., pp. 103–112.

West, R. (2019), The IDEAL Cracking Test, presentation at 2019 NRRA Workshop, https://www.dot.state.mn.us/mnroad/nrra/pavementconference/index.html.

West, R. (2020), A Roadmap to Implementation of Performance Tests in Asphalt Specifications, presented at the 1st FHWA Technical Feedback Group, Oct. 21, 2020.

West, R., C. Rodezno, F. Leiva, and F. Yin (2018). Development of a Framework for Balanced Mix Design, Final report of NCHRP 20-07/Task 406, National Center for Asphalt Technology, August 2018.

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

Williams, R. C., Hill, D., Zelenock, M., and Bausano, J. (2004). Development of Laboratory Performance Test Procedures and Trial Specifications for Hot Mix Asphalt: Final Report. RC-1410. Michigan Department of Transportation.

Yin, F., and West, R. (2021). Balanced Mix Design Resource Guide. National Asphalt Pavement Association. Greenbelt, MD.

Zhou, F., D. Newcomb, C. Gurganus, S. Banihashemrad, et al. (2016), Experimental Design for Field Validation of Laboratory Tests to Assess Cracking Resistance of Asphalt Mixtures, National Cooperative Highway Research Program Project 9-57 Draft Final Report, Transportation Research Board.

Zhou, F., H. Li, S. Hu, et al. (2013), Characterization and Best Use of Recycled Asphalt Shingles in Hot-Mix Asphalt, FHWA/TX-13/0-6614-2, Texas Transportation Institute, College Station, Texas.

Zhou, F., Hu, S., and Newcomb, D. (2020). Development of a Performance-Related Framework for Production Quality Control with Ideal Cracking and Rutting Tests. Construction and Building Materials. 261 (2020). DOI: 10.1016/j.conbuildmat.2020.120549

Zhou, F., Hu, S., Scullion, T., Mikhail, M., and Walubita, L. F. (2007). A Balanced HMA Mix Design Procedure for Overlays. Journal of Association of Asphalt Paving Technology (AAPT). Vol. 76, pp 823-850.

Zhou, F., S. Hu, and D. Newcomb (2020), Development of a performance-related framework for production quality control with ideal cracking and rutting tests, Construction and Building Materials, Volume 261.

Zhou, F., S. Hu, G. Das, and T. Scullion (2011), High RAP Mixes Design Methodology with Balanced Performance, FHWA/TX-11/0-6092-2, Texas Transportation Institute, College Station, Texas.

Zhou, F., S. Hu, G. Das, R. Lee, T. Scullion, and G. Claros (2012). Successful High RAP Mixes Designed with Balanced Rutting and Cracking Requirements, Journal of Association of Asphalt Paving Technologists, Vol. 81, Austin, TX, 2012, pp.477-505.

Zhou, F., S. Hu, T. Scullion, and R. Lee (2014), Balanced RAP/RAS Mix Design System for Project-Specific Service Conditions, Journal of the Association of Asphalt Paving Technologists, Vol. 83.

Zhou, F., S. Hu, T. Scullion, et al. (2007), A Balanced HMA Mix Design Procedure for Overlays, Journal of the Association of Asphalt Paving Technologists (AAPT), Vol. 76, pp. 823–850.

Zhou, F., S. Im, L. Sun, and T. Scullion (2017), Development of an IDEAL Cracking Test for Asphalt Mixture Design, Quality Control and Quality Assurance, Journal of Association of Asphalt Paving Technologists, Newport Beach, California, March 2017.

Zhou, F., Steger, R., and Mogawer, W. (2021). Development of a Coherent Framework for Balanced Mix Design and Production Quality Control and Quality Acceptance. Construction and Building Materials. 287 (2021). DOI: 10.1016/j.conbuildmat.2021.123020

Zhou, F., W. Crockford, and J. Zhang, (2019), Development of an IDEAL Rutting Test for Asphalt Mixture Design, Quality Control and Quality Assurance, Journal of Association of Asphalt Paving Technologists, Fort Worth, Texas.