NCHRP 10-107

National Center for Asphalt Technology
NCAT
at AUBURN UNIVERSITY

Guide for Implementing Balanced Mix Design Specifications

ncat.us

Guide for Implementing Balanced Mix Design Specifications

National Center for Asphalt Technology

April 2023

Contents

Chapter 1. Motivations and Benefits of BMD and Performance Specifications

Objectives:

1) To briefly describe performance specifications utilizing new asphalt mixture performance tests as part of mix design development and approval as well as for Quality Assurance testing.
2) To explain possible motivations for this change including understanding the limitations of the volumetric properties and other traditional asphalt mixture criteria.

Resources:

- Seven case study reports and a summary report by Hajj et al (2021) on the implementation of BMD and performance tests in early adopter states: California, Illinois, Louisiana, Maine, New Jersey, Texas, and Virginia
- National Asphalt Pavement Association's online BMD Resource Guide.

Outcome: Understanding why changes in mix design and acceptance testing are needed and the benefits that can be expected with implementing performance specifications.

1.1 What are Performance Specifications?

The quality of highway materials and construction translates directly to performance and service life; the better the quality, the better the performance and greater the service life of the infrastructure product. By requiring and assessing quality before, during, and following construction (Figure 1.1), performance specifications help agencies incentivize good quality while allowing contractors to be innovative. Agencies may also recover loss of service life by reducing payment for marginal quality in a more rational manner. In that sense, performance specifications are "Next Generation" quality assurance (QA) specifications and represent a progression toward quality measures that are more indicative of how the finished product will perform over time.

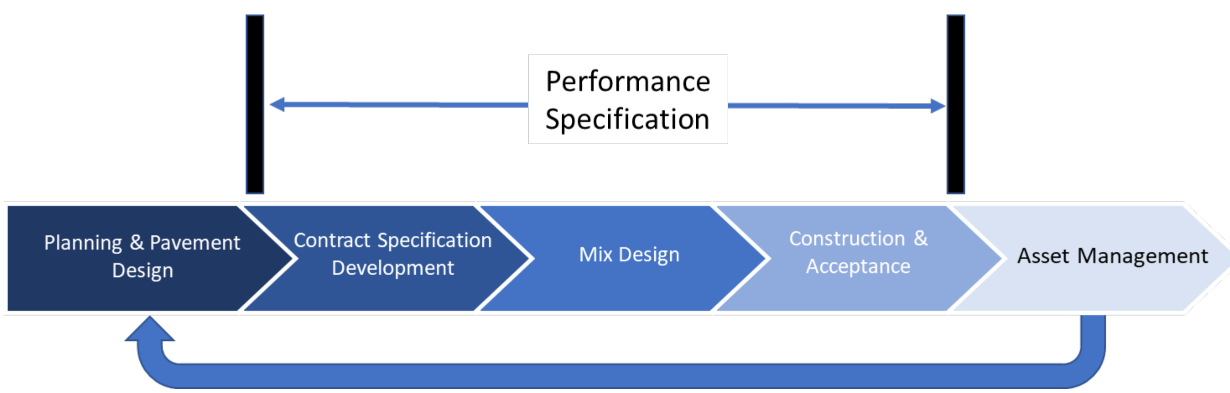


Figure 1.1. Performance specifications as a bridge in the continuum of performance from planning and design to construction and acceptance to asset management.

Performance specification is defined as specifications that described how the finished product should perform over time (E-C235, 2018). It is a broad term used for both asphalt and concrete pavements and encompasses both materials (e.g. balanced mixture design (BMD) for asphalt, performance engineered mixtures (PEM) for concrete) and specifications (performance-related specifications, performance-based specifications, etc.). However, a common thread is the link between the tested quality characteristics and the relevant field performance measures. That link may be an empirical relationship between an index test result and a type of pavement distress or a mechanics-based model that uses fundamental engineering properties to predict damage to the pavement structure over time.

BMD originated as a set of mix performance tests for mix design only. It was later realized that those tests and criteria should be extended to acceptance in order to ensure that the BMD mixture properties were achieved during mix production. Thus, using BMD tests as part of quality assurance, or more specifically, for acceptance of asphalt mixtures is considered a performance specification.

The potential benefits of implementing performance specifications at an agency include the following:

- Performance specifications provide a direct linkage between design expectations and materials and construction quality.
- Performance specifications transfer some of the product performance risk from the agency to the contractor in return for limiting prescriptive requirements, which allows the contractor to have more control and to be more innovative and competitive. In the long run, this will result in better and more efficient construction practices, provide improvements to pavement life, and thus reduce long-term costs to the agency and traveling public.
- Performance specifications often provide incentives for reduced variability. For contractors to succeed using performance specifications and maximize profits, they need to focus on good quality control practices, exceed target quality, and improve consistency.

1.2 Why Change?

The motives for any change are typically rooted in dissatisfaction with the status quo. For example, the status quo could be the agency's current performance period for asphalt overlays, or it could be the limitations that current specifications have on the utilization of more sustainable materials. It is likely that each agency will have several reasons for its interests in implementing BMD and performance specifications.

As noted above, the common thread through BMD and performance specifications is the utilization of tests that provide quality characteristics that are linked to performance. The original vision of the Superpave mix design system was to include performance tests in mix designs for moderate and heavy trafficked pavements. However, the proposed mixture performance tests from the SHRP program were never implemented except for a few special projects, primarily because the tests were not practical for routine use for the thousands of mix designs used each year in the United States.

In the early years of Superpave implementation, most attention was focused on rutting resistance. Improvements in rutting resistance of asphalt mixtures were made by requiring angular aggregates, binder grade adjustments, and higher laboratory compactive efforts. Many highway agencies also added the Asphalt Pavement Analyzer rutting test as a requirement for moderate and heavy trafficked projects. Over the decades since the initial implementation of Superpave, most highway agencies have recognized that rutting has been virtually eliminated, but cracking and other durability distresses have now become a more significant issue. Although there are several factors that may contribute to increased cracking, such as failure to adequately address underlying pavement distresses and problems with construction quality, many asphalt technologists have recognized that changes were needed to increase the asphalt contents of mixtures and limit certain components or additives. Therefore, many agencies have adjusted their mix design requirements and added new binder quality requirements (Tran et al, 2019).

Despite these modifications and limitations, there are still recognized deficiencies in the dependence on volumetric properties for asphalt mix design and QA testing. The two principal volumetric properties currently being used in asphalt mix design and QA specifications are air voids (V_a) and voids in the mineral aggregate (VMA). V_a represents the volume of voids space within a compacted specimen at a specific number of design gyrations (N_{design}) and is widely used as a general indicator of mix quality and a parameter that historically related to rutting resistance. Some agencies also require a minimum VMA for mixture acceptance. The numerical difference between VMA and V_a is the volume of effective binder (V_{be}) in the mix. V_{be} is defined as the total volume of asphalt binder minus the volume of asphalt binder absorbed into the aggregate. In general, a higher V_{be} is desired for better durability and cracking resistance. However, solely relying on V_a and VMA (or V_{be}) has limitations because these parameters provide no indication about the quality of virgin and/or recycled asphalt binders or their interactions with different types of asphalt additives if used. As a result, volumetric properties alone are insufficient to determine how binders (modified or unmodified), recycled materials, and other additives may affect the resistance of a mixture to the range of pavement distresses that highway agencies encounter. For example, consider a given mix design with an unmodified binder and a polymer modified binder. The two mixes may have nearly identical volumetric properties even though the resistance of the mixture containing the polymer modified binder would be expected to have superior rutting and cracking resistance.

Another limitation of volumetric properties is the dependence on aggregate bulk specific gravity (G_{sb}), which is not a reliable property. Some aggregate sources may have consistent G_{sb} values over decades, whereas others have significant variations within a single year due to the site's geology and mining operations. If the G_{sb} values are subject to change over time but are not often verified, the resultant mix designs will have inaccurate volumetric properties. Furthermore, the allowable differences in G_{sb} from lab to lab can lead to significant disputes about the volumetric properties of a mixture. Even relatively small differences in G_{sb} that are well within the allowable range of two results (d2s) of multi-laboratory precision estimates of AASHTO T 84 and T 85 can result in a considerable changes in the calculated VMA and possibly affect the mix design and/or production acceptance decisions. Finally, there is considerable debate as to what is the most accurate method of measuring G_{sb} of aggregates in RAP and RAS. Although different test methods have been adopted by state highway agencies, they do not always yield consistent results and their accuracy varies greatly depending upon the type of aggregate used (Hajj et al 2012).

Lastly, volumetric properties do not provide any indication of the impacts of innovative additives such as fibers, recycling agents, dry-process ground tire rubber, recycled plastics, graphene, carbon nanotubes, etc. on the rutting resistance or cracking resistance of mixtures. Until we can implement a mix design system and acceptance program than can assess true performance related properties for day-to-day projects, innovations will be stymied. In order to identify innovative additives that do and don't improve performance, it is essential that we put the right tools in the hands of agency personnel and contractors rather than asphalt researchers.

In summary, increasing concerns about the durability and cracking issues of asphalt pavements along with a growing awareness of the shortcomings of volumetric criteria for mix design and QA have motivated many highway agencies and the asphalt pavement industry to explore the use of BMD as a new approach to asphalt mix design and production acceptance.

1.3 Benefits of Using Balanced Mix Design

BMD includes the use of performance tests that will provide better assessments of a mixture's resistance to the common forms of pavement distress than volumetric properties. As will be described in more detail later, the BMD tests can be used to *supplement* volumetric criteria or to *replace* some or all volumetric criteria for mix design and QA. Given the known limitations and shortcomings of volumetric criteria, the goal of BMD implementation should be to eventually replace those legacy criteria with performance test criteria. Although some people may be reluctant to relax existing volumetric criteria initially, they should realize that in many cases, significantly higher asphalt contents are needed to improve cracking resistance and durability. Generally, it is not possible to substantially increase asphalt contents without lowering air void targets and raising V_{be} or voids filled with asphalt (VFA) limits. Other legacy criteria such as $N_{initial}$ (Prowell and Brown, 2007) and film thickness (West, 2017) should also be eliminated. Agencies can select specific BMD tests that they are confident to be good indicators of performance and are suitable for routine use in mix design and QA testing.

Typically, a highway agency should select a BMD rutting test and one or more cracking tests depending on the climate, the pavement structure, and underlying layers. Agencies may also require an additional test to evaluate a mixture's resistance to moisture damage. A summary of BMD tests can be found on NAPA's online BMD Resource Guide.

Most BMD tests have been shown to be sensitive to mix component materials such as asphalt content and asphalt grade, recycled materials contents, and the use of novel mixture additives such as recycling agents and fibers. Thus, mix designers can better optimize mixtures to meet an agency's BMD criteria with sustainability interests in mind.

Furthermore, BMD criteria can be adjusted for specific pavement applications and layers of a pavement structure that considers the expected traffic (loading) conditions, aging gradient with depth, and underlying conditions of the existing pavement for overlay applications.

One note of caution is that implementing BMD should not be expected to remedy issues with underlying layers in a pavement. Although BMD mixtures can be expected to have better cracking resistance than many currently used asphalt mixtures, they will not solve problems with reflection cracking, or lower layers damaged by debonding, stripping, or fatigue cracking, as illustrated in Figure 1.2. Those issues must be addressed through proper rehabilitation designs and strategies at the project and agency level.

(a) Reflection cracking in an overlay from the underlying concrete pavement	(b) Core from a pavement with a stripped intermediate layer	(c) Core showing an incompetent leveling mix with debonding

Figure 1.2. Examples of Pavement Problems that BMD Will Not Fix

It is also important to realize that the path to full BMD implementation is not completely paved at this time. Some research and information gaps still must be addressed and new or modified processes need to be developed as we go along. For example, there is a lack of guidance for state DOTs in selecting among the multiple mix design approaches in AASHTO PP 105 and the many mixture performance tests for the implementation of BMD. Furthermore, questions like ‘how to address asphalt aging and its impact on mixture performance tests during mix design and production’ and ‘how to incorporate quality assurance into a BMD system’ remain to be answered. However, it is time to begin planning the process and work on the missing details.

Table 1.1 summarizes benefits of BMD cited by early adopters among state highway agencies. Many of these highway agencies began using mix performance tests for mix design approval of certain mix types and then progressed toward broader implementation for mix design and field acceptance testing over time. Most of the highway agencies have adopted these performance requirements to extend the lives of their asphalt pavements or to optimize the use of RAP and other recycled materials without sacrificing pavement performance. Other expected benefits of BMD testing cited by the asphalt pavement construction industry include the ability to better assess the use of innovative asphalt additives, relaxing or eliminating volumetric requirements, and more robust methods for mix design approval and production acceptance (Yin and West, 2021).

Table 1.1 Early state DOT adopters of BMD, their target pavement applications, benefits and contacts.

State DOT	Applications	Performance Tests Used	Expected Benefits	DOT Contacts
New Jersey	Specialty mixes such as High RAP, High Performance Thin Overlays (HPTO), and Binder Rich Intermediate Course (BRIC)	Rutting: APA (AASHTO T 340) Cracking: BBF (AASHTO T 321), OT (NJDOT B-10)	Overall improvement in the condition of the state DOT pavement network by extending the life of asphalt pavements	Stevenson Ganthier Mark Gillece Rajesh Kabaria Ryan Rathbun Robert Blight
California	Long Life Asphalt Pavement (LLAP) projects	Rutting: FN (Modified AASHTO T 378), HWTT (AASHTO T 324) Cracking: BBF (AASHTO T 321), I-FIT (AASHTO T 393)	Optimized use of recycled materials without jeopardizing pavement performance	Thomas Pyle Kee Foo Raghubar Shrestha Pete Spector
Illinois	Low and High ESAL mixes and SMA	Rutting: HWTT (Modified AASHTO T 324) Cracking: I-FIT (Modified AASHTO T 393)	Enhanced the quality of asphalt mixtures and allow Contractors to optimize the use of recycled materials	Brian Pfeifer Jim Trepanier Tom Zehr Brian Hill Greg Heckel
Louisiana	All asphalt mixtures	Rutting: HWTT (AASHTO T 324) Cracking: SCB (DOTD TR 330)	Design and production of more economical, higher RAP content mixes without jeopardizing pavement performance	Luanna Cambus Jason Davis Sam Cooper III Corey Mayeux
Texas	Initially used for certain specialty mixes; Pilot BMD projects started in 2019	Rutting: HWTT (Tex-F-242) Cracking: OT (Tex-F-248)	Production of cost-effective asphalt mixtures that meet the project performance specifications	Ryan Barborak Travis Patton Enad Mahmoud
Virginia	Initially used to evaluate higher RAP content mix designs and field trial	Rutting: APA (AASHTO T 340) Cracking: IDEAL-CT (ASTM D8225), Cantabro (AASHTO TP 108)	Better utilization of recycled materials and innovative additives to resist the common forms of distress	Kevin McGhee Stacey Diefenderfer Ilker Boz Jhony Habbouche

1.4 The Steps in the Process of Implementing Performance Specifications

People that were in this field 20 to 25 years ago remember the implementation of Superpave binder and mixture specifications. In many states, full implementation was accomplished within a matter of a few years, including purchasing of new equipment, training people on the new standards, adopting new asphalt binder and mix design criteria and making major changes to quality assurance programs. Over the subsequent years, many of the initial Superpave requirements were revised as we learned what aspects were unnecessarily challenging and inconsequential. Reflecting on that experience is useful to better prepare for the next big change in specifications for asphalt materials.

There are numerous steps or tasks in the process to implement changes as significant as those involved in deploying new performance specifications for mix design approval and acceptance testing. Figure 1.3 illustrates the major steps that are described later in this guide. Overall, this process may take five years or more to complete, depending on the goals, how the effort is planned, and the resources made available to the process.

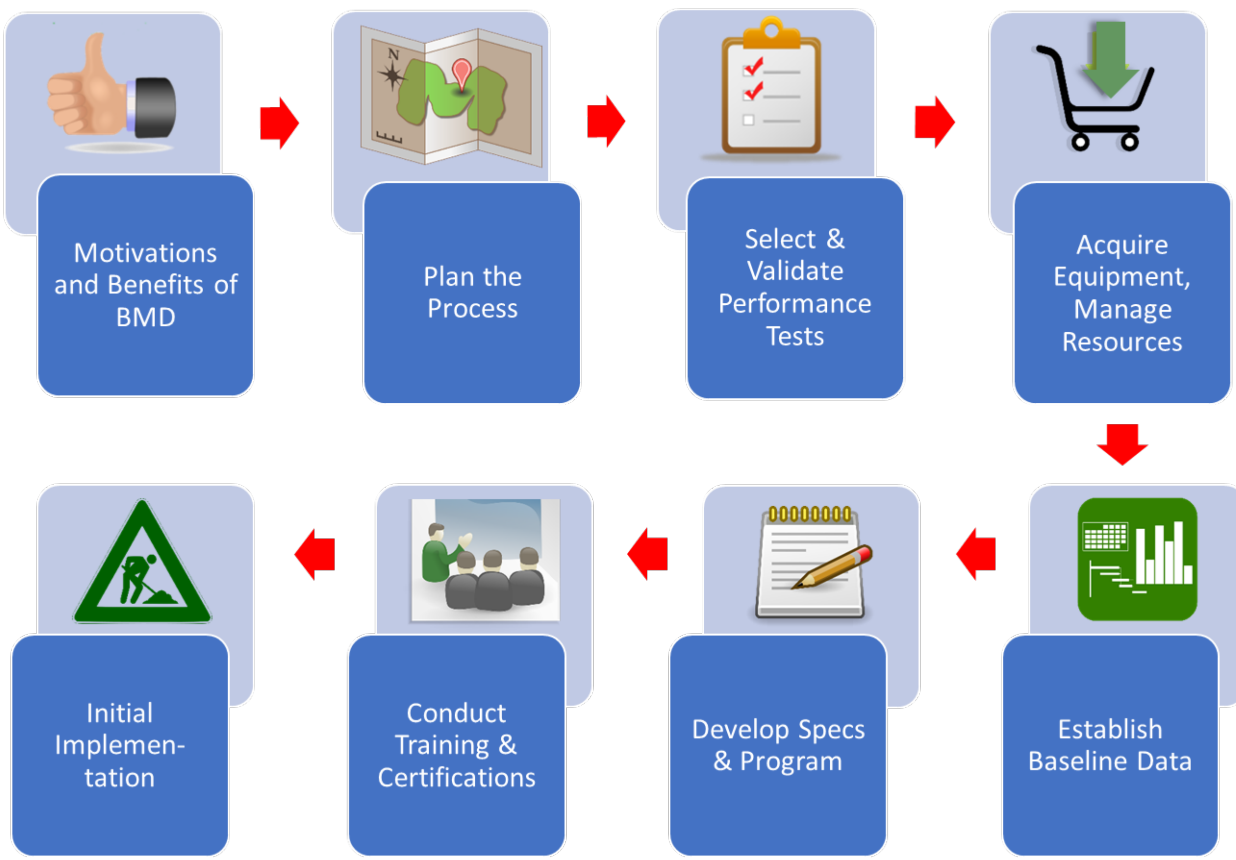


Figure 1.3. Illustration of the Major Steps in the Process of Implementing Performance Specifications.

References

Glossary of Transportation Construction Quality Assurance Terms, Transportation Research Circular Number E-C235, 7th Edition (2018) Transportation Research Board.

Hajj, E. Y., P. E. Sebaaly, R. West, N. Morian, and L. Loria (2012) Recommendations for the Characterization of RAP Aggregate Properties Using Traditional Testing and Mixture Volumetrics. *Road Materials and Pavement Design*, Vol. 13, No. S1, pp. 209-233.

Hajj, E. Y., T. B. Aschenbrener, Nener-Plante, D, (2021) Case Studies on the Implementation of Balanced Mix Design and Performance Tests for Asphalt Mixtures: California Department of Transportation (Caltrans), WRSC-TR-21-02.

Hajj, E. Y., T. B. Aschenbrener, Nener-Plante, D, (2021) Case Studies on the Implementation of Balanced Mix Design and Performance Tests for Asphalt Mixtures: Illinois Department of Transportation (IDOT), WRSC-TR-21-03.

Hajj, E. Y., T. B. Aschenbrener, Nener-Plante, D, (2021) Case Studies on the Implementation of Balanced Mix Design and Performance Tests for Asphalt Mixtures: Louisiana Department of Transportation & Development (LaDOTD), WRSC-TR-21-04.

Hajj, E. Y., T. B. Aschenbrener, Nener-Plante, D, (2021) Case Studies on the Implementation of Balanced Mix Design and Performance Tests for Asphalt Mixtures: Maine Department of Transportation (MaineDOT), WRSC-TR-21-05.

Hajj, E. Y., T. B. Aschenbrener, Nener-Plante, D, (2021) Case Studies on the Implementation of Balanced Mix Design and Performance Tests for Asphalt Mixtures: New Jersey Department of Transportation (NJDOT), WRSC-TR-21-06.

Hajj, E. Y., T. B. Aschenbrener, Nener-Plante, D, (2021) Case Studies on the Implementation of Balanced Mix Design and Performance Tests for Asphalt Mixtures: Texas Department of Transportation (TxDOT), WRSC-TR-21-07.

Hajj, E. Y., T. B. Aschenbrener, Nener-Plante, D, (2021) Case Studies on the Implementation of Balanced Mix Design and Performance Tests for Asphalt Mixtures: Virginia Department of Transportation (VDOT), WRSC-TR-21-08.

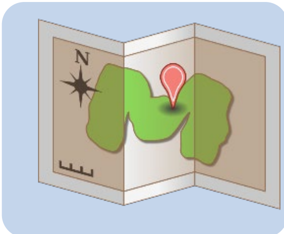
Hajj, E. Y., T. B. Aschenbrener, Nener-Plante, D, (2021) Positive Practices, Lessons, Learned, and Challenged when Implementing Balanced Design of Asphalt Mixtures: Site Visits, WRSC-TR-21-01.

Prowell B.D. and E.R. Brown (2007) Superpave Mix Design: Verifying Gyration Levels in the N_{design} Table, National Academies of Sciences, Engineering, and Medicine 2007. Superpave Mix Design: Verifying Gyration Levels in the Ndesign Table. Washington, DC: The National Academies Press. https://doi.org/10.17226/23207.

Tran, N., G. Huber, F. Leiva, B. Pine, and F. Yin (2019) Mix Design Strategies for Improving Asphalt Mixture Performance, NCAT Report 19-08.

West, R. (2017) Asphalt Film Thickness Debunked, Asphalt Technology News, National Center for Asphalt Technology, Vol. 29, No. 1.

Yin, F., and R. West (2021) Balanced Mix Design Resource Guide, NAPA Information Series 143.



Chapter 2. Overall Planning

Objectives:

1) To understand the overall implementation process,
2) to define and set state DOT goals for implementation, and
3) to determine the resources necessary to achieve those goals within a realistic timeline.

Resources:

- List of Major Tasks and Subtasks to Implement Performance Specifications, see Table 2.4
- Example Gantt Chart for BMD Implementation

Outcome: Understanding the full implementation process as well as the goals, required resources, and timeline needed for implementation.

2.1 Identification of Champions

Having champions in the state DOT and industry to provide leadership for the various implementation activities, as well as helping to overcome institutional roadblocks and technical hurdles will be critical in successfully implementing performance specifications. For a state DOT, this involves continuous communication and internal partnerships among various offices, such as materials, construction, pavement design, and pavement management. Other areas critical to implementation include gaining support from industry, as well as support and potential resources from the local FHWA Division Office. Leadership from the asphalt contracting industry may be found in industry associations and/or individual paving contractors and suppliers.

When identifying a champion, an ideal candidate would be someone at a level within the organization that has the authority necessary to overcome potential agency-related roadblocks; has a thorough understanding of the implementation subject; and has the time to dedicate to the implementation effort. While someone at a Chief Engineer's level would certainly have the authority, it is doubtful they would have the time to devote to the effort or have the background to understand many of the details that will be affected by the change. One approach that has been used successfully by a number of states is to appoint a champion, typically from the central office or central lab, who has a good understanding of the technical issues, and also good leadership and communication abilities. While this may be someone at the level of a State Construction Engineer or State Materials Engineer, the ideal candidate would likely be one level lower, such as a State Bituminous Engineer or a Deputy State Construction Engineer. The key for this approach to be successful is that the candidate must be given adequate authority and support, so they are capable of dealing with impediments to implementation.

2.2 Establishing a Stakeholders Partnership

One key element of successfully developing and implementing performance specifications is the formation of a joint Agency/Industry Task Force within the state to provide higher-level stakeholder input in the numerous decisions along the path towards implementation, and to make certain that critical issues are properly addressed. The Task Force is also an excellent method of assuring that there is continual support and buy-in from both industry and agency. Members of the Task Force would potentially consist of representatives from the following stakeholder groups:

- Implementation Champion from the Agency (Chair)
- Asphalt Pavement Association President/Technical Director (Co-Chair)
- Company President/General Manager (3 – 4)
- FHWA representative (1)
- DOT District Materials and/or Construction representatives (2 – 3)
- DOT Central Office Materials and Construction representative (2 – 3)
- Academia representative (1 – 2)

The Champion, along with the Task Force would provide the “out front” leadership role and would be available to provide periodic status updates for the initiative to executive stakeholders, and a periodic forum for dialogue about progress and key milestones as the effort progresses.

In addition to a higher-level Task Force, a number of states have also found that creating a Technical Committee (i.e. a BMD Implementation Committee) is necessary to address many of the detail-oriented tasks involved in implementing a new program. The Technical Committee would report to the Task Force and regularly oversee the progress of related projects and activities, provide technical input, and support the development of the performance specifications, procedures, training programs, and guidance. The Technical Committee would typically consist of representation from the DOT’s Central Office (Materials, Construction, and Pavement Management), as well as District representatives with hands-on project experience. Industry representatives would consist of experienced QC Managers, Mix Designers, and Project Superintendents.

One essential component of both the higher-level Task Force as well as the detail-oriented Technical Committee is that the meetings and decision making should be well documented. This will provide valuable assistance in maintaining a consistent understanding of the direction of the implementation and will also provide continuity in situations where there is a change of leadership.

Establishing or having an established partnership between a state DOT and academia also can go a long way in supporting critical and pressing research that may be needed as part of the development and implementation efforts. In many instances, problems arise that can be addressed in small-scale research projects, conducted in a timely manner either in-house by the agency or by academia. Examples of these types of partnerships are shown in Table 2.1.

States with limited resources may find it helpful to work with other states in the region, particularly in areas with similar environments, materials, and contractors. This would expand the overall experience level, allow for the sharing of data, increase participation in round-robin studies (discussed later) and help to address problems common to each state.

Table 2.1. Examples of Partnerships between State DOTs and Academia.

State DOT	Academia Partnerships
Virginia	The Virginia Transportation Research Council (VTRC) is a partnership of the VDOT and the University of Virginia and is responsible for all research at VDOT.
New Jersey	Rutgers University's Center for Advanced Infrastructure and Transportation (CAIT) Pavement Support Program (PSP) is funded by the State Planning and Research (SP&R) Program and is responsible for providing pavement engineering support to the NJDOT's Pavement and Drainage Management Systems (P&DMS) Unit.
Texas	Texas DOT utilizes several academic institutions to assist with research aimed toward implementing asphalt performance specifications including the Texas A&M Transportation Institute, the University of Texas (Austin), the University of Texas El Paso, and the National Center for Asphalt Technology (NCAT) at Auburn University.
Wisconsin	The Wisconsin Highway Research Program contracts research projects to organizations through competitive Requests for Proposals. Research organizations that have contributed to Wisconsin's BMD implementation efforts include the University of Wisconsin (Madison), Advanced Asphalt Technologies LLC, and NCAT.

2.3 Doing Your Homework (Identifying the Issues, Identifying Resources, and Reviewing Literature)

One of the first steps involved in implementing asphalt performance specifications is developing the appropriate background historical information necessary for implementation. This could include information such as:

- What are the concerns and performance problems with existing asphalt mixtures and specifications?
- What types of resources will be available for the implementation effort, both financially and intellectually?
- How have other implementation efforts in the state fared? What worked well and what didn't?
- What have other agencies experienced when they have adopted similar technologies? Is there available literature that can be reviewed?

Compiling this type of background information will go a long way in getting the implementation effort off on the right foot and will set the stage for further efforts.

Identifying concerns and performance problems with existing asphalt mixtures and specifications will help ensure that implementation efforts are ultimately targeting the right areas. Examples of the types of problems that could be addressed by performance specifications include premature cracking due to under-asphalted mixes, high mixture costs due to the lack of locally available material sources, or overly restrictive specifications that limit the use of recycled materials or don't take into consideration new and innovative materials and additives. Once the problem areas are identified, the agency needs to consider how performance specifications can help to resolve them. Examples of state DOTs that have targeted specific issues being addressed with performance specifications are shown in Table 2.2.

Table 2.2. Examples of Key Issues Being Addressed by State DOTs.

State DOT	Issue Addressed
California DOT (Caltrans)	Performance of high-traffic asphalt pavements
Illinois DOT (IDOT), Louisiana DOT and Development (LADOTD), VDOT	Recycled materials
Maine DOT	Premature failure of asphalt mixtures
NJDOT and TxDOT	High-performance and specialty asphalt mixtures

A commitment to providing resources to implement performance specifications is needed from senior leadership and management of a state DOT. Thus, it is important for a state DOT to identify its available resources and funding levels that can be devoted to the development and implementation of BMD process. A high-level assessment of current organizational structure, staffing and laboratory readiness levels, workspace availability, and annual asphalt tonnage, is needed. In addition, an assessment must also be made on the potential impact to contractors and their operations. A more detailed analysis can be conducted as described in Task 4.

A state DOT also needs to conduct a critical review of available literature related to performance specifications and gather information on completed or undergoing implementation efforts and activities by other state DOTs. The general purpose of this is to avoid reinventing the wheel, so to speak. This includes reviewing related research projects and findings, specifications, and procedures, participating in related workshops; and meeting with other state DOTs to learn about their experiences with performance testing and specifications. It is also important for a state DOT to understand the factors that drive asphalt mixture performance, including issues with existing rehabilitation strategies and construction practices that need to be resolved before or concurrent to implementing performance specifications. For example, implementing new mixture performance requirements without addressing pavements with poor underlying layers could result in in the performance requirements taking the blame for early projects that perform poorly.

Having the means to conduct peer exchanges of relevant information, positive practices, challenges, and lessons learned is a practical and effective approach to foster communication and coordination among state DOTs for successful development and implementation of BMD and performance specifications.

2.4 Establishing Goals

While recognizing the overall benefits from Task 1, a state DOT, in partnership with stakeholders, needs to establish its state-specific goals for its performance specification program to help guide decision making. The goals are usually specific to each state considering its organizational structure, staffing levels, workspaces, budget, annual asphalt tonnage, as well as industry experiences and practices. State DOT goals related to asphalt performance specifications can focus on several areas, such as:

- Overall improvement of asphalt pavement performance,
- Optimization of recycled materials usage,
- Ability to target specific pavement layers (e.g., surface, intermediate, base),
- Development and utilization of specialty asphalt mixtures for specific pavement applications,
- Improving acceptance practices, and/or
- The design and construction of long-lasting asphalt pavements for high investment projects.

Thus, the overall scope of the performance specification program needs to be defined. Since state DOTs have asphalt programs of different sizes, the plan for implementation may need to consider factors related to the scope of the project, such as project investment level (e.g., high investment paving projects), full reconstruction versus major rehabilitation (e.g., full depth reclamation) or minor rehabilitation (e.g. mill and overlay), highway functional classification (principal arterial, minor arterial, etc.) and project traffic level (low, moderate, or high volume), project length and asphalt tonnage; pavement layer (e.g., surface, binder, base course), outcome of a risk-based, cost-benefit analysis for the project; etc. For instance, a range of examples of BMD project scopes are shown in Table 2.3. Some states have decided to implement BMD for all projects, whereas others have initially limited BMD specifications to certain project or mixture types.

Some states may also consider a phased approach to implementation. Some may plan to begin with implementing BMD mix design specifications first and work toward implementing the performance tests as part of QA in a year or two. An advantage of this approach is that it gives the agency and contractors the opportunity to get familiar with the tests and their materials in a limited setting before adding the complications of implementing the tests and criteria in the field. A disadvantage of this approach is that it assumes that the current tests, specifications, and practices for QA provide a satisfactory assessment of mixture quality. That may not be the case as explained in Subtask 1.2. Therefore, to realize the complete benefits of BMD, the mix performance tests must be implemented for both mix design approval and acceptance.

Since the implementation of performance specifications will take several years to be fully completed, it is essential that the overall goals and program scope be well communicated. Over time, people change positions, as does management. Having the direction for the process in writing helps to provide continuity in a time when things change. One thing to note is that throughout the overall implementation effort, the identified goals and program scope may need to be adjusted as the state DOT and contractors gain experience with the performance specification process.

Table 2.3. Examples of Balanced Mix Design Program Scopes Considered by State DOTs.

State DOT	Scope
Caltrans	High-traffic paving projects with more than 100,000 tons of asphalt mixture produced.
MaineDOT	Interstate and high investment paving projects.
LADOTD, IDOT, and NJDOT	All projects for asphalt mixture design.
TxDOT	Initially, certain specialty asphalt mixtures; but currently, all dense-graded asphalt mixtures.

2.5 Mapping the Tasks

The major tasks and subtasks to successfully implement performance specifications as part of a state's mix design approval process and QA program are listed in Table 2.4. Further details regarding each of the subtasks are provided in this guide. A state may find some of these tasks are not applicable for their situation and/or goals. Further, a state may identify additional tasks or subtasks that are necessary. It should be noted that some tasks identified in this guide may overlap or be conducted concurrently.

Although a state DOT may have already completed some tasks and sub-tasks, it is important to periodically review earlier tasks to make sure that all issues have been addressed. Skipping tasks may seem like a good idea to move the process along quickly, but eventually the overlooked task will be evident and may require revisiting earlier tasks that were thought to have been completed. Additionally, it should be noted that in developing the overall implementation plan it is important to have discussions of the tasks with the partnership of stakeholders to decide how to address each task and sub-task. Some tasks may be best handled in a cooperative effort among several states in a region.

Table 2.4. Suggested Major Tasks and Subtasks to Implement Performance Specifications.

Task		Sub Task	Description
1	Motivations and benefits of BMD and Performance Specifications		
2	Overall Planning	2.1	Identification of Champions
		2.2	Establishing a Stakeholders Partnership
		2.3	Doing Your Homework
		2.4	Establishing Goals
		2.5	Mapping the Tasks
		2.6	Developing an Implementation Timeline
		2.7	Identifying Available External Technical Information and Support (periodically)
3	Selecting and Validating Performance Tests	3.1	Identifying Primary Modes of Distress
		3.2	Identifying and Assessing Performance Test Appropriateness
		3.3	Dealing with Reheating and Conditioning/Aging of Mixtures
		3.4	Relationship Confirmation and Criteria Development
4	Performance Testing Equipment and Staffing: Acquiring, Managing Resources, Training, and Evaluating	4.1	Acquiring Equipment
		4.2	Managing Resources
		4.3	Conducting Initial Training
		4.4	Refining Procedures
		4.5	Conducting Interlaboratory Studies
5	Establishing Baseline Data	5.1	Reviewing Existing Lab Results & Pavement Management System Data
		5.2	Conducting Benchmarking Studies
		5.3	Conducting Shadow Projects
		5.4	Analyzing Production Data
		5.5	Determining How to Adjust Mixtures Containing Local Materials
6	Specifications and Program Development	6.1	Sampling and Testing Plans
		6.2	Pay Adjustment Factors (If Part of the Goals)
		6.3	Developing Pilot Specifications and Policies
		6.4	Conducting Pilot Projects
		6.5	Final Analysis and Specification Revisions
7	Training, Certifications, and Accreditations	7.1	Developing and/or Updating Training and Certification Programs
		7.2	Establishing or Updating Laboratory Accreditation Program Requirements
8	Initial Implementation		

2.6 Developing an Implementation Timeline

Establishing an implementation timeline is important to guide the goals (Subtask 2.4) toward a successful and timely completion within the available resources (Subtask 2.3). The scope (also in Subtask 2.4) plays an important role in the timeline as well. Implementation can be accomplished in phases.

An implementation plan needs to be established and broken down into tasks and activities that have their own segmented time frame. The implementation plan involves using the tasks and subtasks identified in this document as a framework. A Gantt chart can be developed for keeping track of these time frames by showing how long each task should take, which tasks can be done at the same time, and which should be done sequentially. The chart will help to monitor the progress of various activities and if a particular task takes longer than planned, it will show whether that delay will affect the time frame of other tasks and ultimately the overall implementation time frame.

An example Excel file Gantt chart for the tasks and subtasks identified in this guide is provided at this link. Table 2.5 provides preliminary guidance on the expected duration of each subtask. The subtasks that are expected to take at least a year to complete are briefly outlined below.

Subtask 3.4 Relationship Confirmation and Criteria Development (60 months)

A critical factor in selecting performance tests is having strong relationships to field performance, and those relationships are useful in setting specification criteria specific to a state's traffic, climate, materials, and existing pavement structures. Very few well-controlled field validation experiments have been conducted to date. To address this gap, field experiments with at least six pavement test sections should be constructed for each of the common distresses that the performance tests are intended to assess. These field validation efforts generally take several years before sufficient pavement performance data is generated that distinguishes good and bad performance. Failing to establish reliable relationships between the laboratory tests and field performance could result in selecting poor tests and/or setting inappropriate specification criteria for the tests.

Subtask 4.1 Acquiring Equipment (up to 24 months)

Depending on the cost of the equipment for the selected tests and the number of test units and ancillary devices required for specimen preparation that need to be purchased by the state and contractors, the time between the decision to buy and actual delivery could be lengthy. Timing could also be affected by budget cycles for organizations. In general, more sophisticated tests will require more expensive equipment and longer lead times. In such cases, purchases could be made in phases so that the more critical labs are equipped first.

Subtask 5.2 Conducting Benchmarking Studies (12 months)

Benchmarking studies provide a primary source of data that can be used to help set appropriate specification limits for performance tests. These studies involve testing samples from existing mix designs and plant produced mixtures that are currently used in the state. It is important for the benchmarking work to include samples representing materials from different parts of the state as well as mixtures representing the state's different traffic levels. Ideally, 15 to 25 mixtures should be included in each traffic category for a suitable analysis. In general, benchmarking studies will take a year or more to complete.

Subtask 5.3 Conducting Shadow Projects (15 months)

Shadow projects are paving projects for which additional plant mixture samples are collected and performance tests are conducted primarily to gather information on production variability. Other benefits of shadow projects are that they help familiarize both agency and contractor personnel with the selected tests, and the results can supplement the benchmarking database. A minimum of 10 shadow projects are recommended covering the different parts of a state, and for each project, tests should be conducted on each sublot for two to three full lots. A few months should be built into the beginning of the subtask to organize the shadow project plan.

Subtask 5.4 Analyzing Production Data (12 months)

Preliminary analysis of the data from the shadow projects should occur as the data are being generated, so this task can essentially run concurrently to Subtask 5.3. The final analysis of the compiled data from the shadow projects may take approximately one additional month.

Subtask 5.5 Determining How to Adjust Asphalt Mixtures Containing Local Materials (15 months)

The purpose of this subtask is to give contractors time to learn how to adjust their mixtures to meet future performance criteria as economically as possible. Thus, this subtask occurs after the contractors have purchased equipment and are familiar with the new test methods, but before the criteria have been established for the pilot projects. This subtask can also run concurrent to the preceding three subtasks.

Subtask 6.4 Conducting Pilot Projects (18 months)

Pilot projects are the first real assessment of the new QA program under actual contractual conditions. The pilot projects start with the typical bidding-contracting process with the new QA requirements and specifications enforced, including performance testing required as part of asphalt mixture design and acceptance. The plan for the necessary pilot projects could begin with just a few projects in the first year, before increasing to include additional districts and contractors the next year. Following each pilot project, agency and contractor personnel should have a close-out meeting to discuss problems and concerns that need to be addressed for future pilot projects and prior to finalizing the program for full implementation.

As previously noted, the overall timeline for full implementation may be more than five years. Many states have already begun work or completed some of the subtasks. Each state can modify the timeline for their own circumstances.

Table 2.5. Preliminary Guidance on Durations and Starting Times for Implementation Tasks and Subtasks.

	Task	Sub Task	Description	Expected Start (Month)	Duration (Months)
1	Motivations and benefits of Performance Specifications			-4	4
2	Overall Planning	2.1	Identification of Champions	1	1
		2.2	Establishing a Stakeholders Partnership	2	1
		2.3	Doing Your Homework	2	2
		2.4	Establishing Goals	3	2
		2.5	Mapping the Tasks	3	3
		2.6	Developing an Implementation Timeline	3	3
		2.7	Identifying Available External Technical Information and Support (periodically)	2	--
3	Selecting and Validating Performance Tests	3.1	Identifying Primary Modes of Distress	6	1
		3.2	Identifying and Assessing Performance Test Appropriateness	7	3
		3.3	Dealing with Reheating and Conditioning/Aging of Mixtures	7	12
		3.4	Relationship Confirmation and Criteria Development	10	60
4	Performance Testing Equipment and Staffing: Acquiring, Managing Resources, Training, and Evaluating	4.1	Acquiring Equipment	10	1 - 24
		4.2	Managing Resources	27	2
		4.3	Conducting Initial Training	28	2
		4.4	Refining Procedures	28	6
		4.5	Conducting Interlaboratory Studies	34	6
5	Establishing Baseline Data	5.1	Reviewing Existing Lab Results & Pavement Management System Data	33	1
		5.2	Conducting Benchmarking Studies	34	12
		5.3	Conducting Shadow Projects	37	12
		5.4	Analyzing Production Data	34	15
		5.5	Determining How to Adjust Mixtures Containing Local Materials	34	15
6	Specifications and Program Development	6.1	Sampling and Testing Plans	46	6
		6.2	Pay Adjustment Factors (If Part of the Goals)	49	3
		6.3	Developing Pilot Specifications and Policies	46	6
		6.4	Conducting Pilot Projects	52	18
		6.5	Final Analysis and Specification Revisions	67	6
7	Training, Certifications, and Accreditations	7.1	Developing and/or Updating Training and Certification Programs	73	6
		7.2	Establishing or Updating Laboratory Accreditation Program Requirements	34	3
8	Initial Implementation			80	--

2.7 Identifying Available External Technical Information and Support

New information is constantly coming out about performance tests and related issues through ongoing research studies and experiences of other states. Many State DOT's have internal research underway or have engaged with their universities or other research organizations to provide information to guide decisions related to implementation of performance tests. Thus, a regular review of related information and observations from other state DOTs is needed throughout the performance specification implementation process. State DOTs may also reach out to the FHWA Resource Center for technical assistance, training, technology deployment, and interagency coordination. Having periodic peer exchanges can help a state DOT to ensure its implementation program remains viable and productive. NCAT has conducted numerous BMD workshops over the past few years to share the latest information. Through these workshops, NCAT has gained insight on the directions and pace that many state DOTs are moving on the topic. The National Asphalt Pavement Association (NAPA) and NCAT have developed the online BMD resource guide that provides up to date information on the status of BMD specification development across the USA.

Chapter 3. Selecting and Validating the Performance Tests

Objectives:

1) To explain the key factors in selecting asphalt mixture performance tests for mix design development, approval, and QA testing, as well as a logical process by which those decisions can be made.
2) Guidance is also provided on developing an experiment to validate performance tests and criteria for a state or local agency.

Resources:

- Checklist of steps for selecting performance tests and establishing performance criteria, see p. 27.
- NAPA online Balanced Mix Design Resource Guide
- FHWA lead pooled fund TPF-5(478) Demonstration to Advance New Pavement Technologies
- FHWA-HIF-19-103 Report: Index-Based Tests for Performance Engineered Mixture Designs for Asphalt Pavements

Outcomes: (1) Identify the primary types of asphalt pavement distress for the state or local highway agency. (2) Identify appropriate tests or methods to address the primary distresses. (3) Plan for validation of the performance tests.

3.1 Identifying the Primary Modes and Causes of Distress

Having a clear understanding of the types of pavement distress frequently encountered by the highway agency is the first step in developing an appropriate plan for how to remedy those issues. For example, if one of the most common pavement distresses is bottom-up fatigue cracking, then the appropriate solutions are likely to be very different than dealing with thermal cracking or rutting.

It should be noted that BMD has the potential to better address pavement distresses that are primarily related to the quality of the material than those caused by non-material related factors. Notable examples of these distresses are top-down cracking, thermal cracking, and rutting. Top-down cracking is a mode of surface cracking distress that has been often encountered by state and local highway agencies. Different from the traditional bottom-up fatigue cracking, top-down cracking initiates at the surface of the asphalt pavement in or near the wheel-path and gradually progresses downward through the asphalt layers. Previous research has shown that top-down cracking develops due to a combination of high bending-induced surface tension and shear-induced near-surface tension from tire-pavement interactions. Factors that have been identified to contribute to the development of top-down cracking include traffic loading, pavement structure, and mixture properties. Specifically, asphalt mixtures with inadequate durability and cracking resistance and those that are highly susceptible to asphalt aging are typically prone to top-down cracking. Therefore, implementing BMD has a great potential to address top-down cracking as it requires characterizing the cracking resistance of asphalt mixtures for mix design approval and/or production acceptance, especially when the impact of asphalt aging is considered when conducting the mixture cracking test for BMD.

For some distresses, such as debonding or reflection cracking, implementing BMD tests alone should not be expected to fix the problems. Debonding, for example, is more likely due to issues with tack coat materials, application rates, and/or paving practices once the material is applied. Fixing reflection cracking, on the other hand, has more to do with addressing problems in the underlying pavement layers than the use of more crack resistant asphalt overlay mixtures since reflection cracking will eventually propagate through the most robust asphalt mixes. Nevertheless, use of crack resistant mixes has the potential of delaying the development of reflection cracking. Bottom-up fatigue cracking is another distress that cannot be addressed by implementing BMD tests alone because its development is primarily due to the tensile strains at the bottom of the asphalt layer exceeding the tensile strength of the asphalt mix. Therefore, having a robust pavement structure from the pavement design perspective is likely to be more effective in mitigating bottom-up fatigue cracking than implementing BMD cracking tests for mix design and production acceptance. Block cracking, on the other hand, is highly related to the quality of asphalt binder (including both virgin binder and recycled binder) in the mix. Unfortunately, there is a lack of information on how the existing BMD cracking tests correlate to block cracking performance of an asphalt pavement. Therefore, implementing BMD tests alone may not be able to address block cracking issues unless further research and development efforts are devoted to developing and validating a robust BMD test for block cracking evaluation.

The primary source of information for this task is the agency's Pavement Management System. Most agencies rate the pavement condition of their highway network every one to three years. From these surveys, the agency can determine the most common forms of distress (e.g., rutting, cracking, roughness) by region, roadway classification, and age since the last rehabilitation. However, many pavement condition surveys do not adequately categorize cracking in such a way that the cause of cracking can be easily inferred. For example, cracking caused by bottom-up fatigue, top-down cracking, moisture damage in asphalt layers, and debonding of layers can all have similar initial appearances. For this reason, a pavement coring program is an essential complementary activity to help understand the type of pavement distresses. Many state highway agencies have a coring program to determine appropriate pavement rehabilitation strategies for pavements at or near their pavement condition thresholds. The agency's regional engineers may have significant experience with pavement performance, maintenance, and rehabilitation and could provide useful insight on the primary modes and causes of pavement distress in their jurisdiction.

Figure 3.1. (Left) Distressed pavement with apparent wheel-path fatigue cracking. (Right) One of several cores taken from the project that showed debonding below the surface layer.

An executive summary report should be prepared for the Technical Committee (i.e., BMD implementation committee) that highlights the state's common pavement distresses for selecting the appropriate mixture test(s) for those issues.

3.2 Identifying and Assessing Performance Test Appropriateness

Identifying candidate performance tests to address the primary pavement distresses should involve agency, industry, and consulting lab perspectives, especially in cases where contractor results are used in acceptance decisions since the tests will affect staffing and equipment purchasing decisions for those labs.

Several possible mixture performance tests can be used to assess an asphalt mixture's resistance to most types of distress. Useful references for the selection of performance tests include AASHTO provisional standards for BMD (i.e., PP 105-20 and MP 46-20), the final report from NCHRP 20-07/Task 406, the NAPA Balanced Mix Design Resource Guide (Yin and West, 2021), and the report by Hajj et al. (2019) titled *Index-Based Tests for Performance Engineered Mixture Designs for Asphalt Pavements*. The NAPA Balanced Mix Design Resource Guide provides a one-page summary for each of the BMD performance tests being considered by state highway agencies and the asphalt pavement industry.

Selection of the tests should consider the agency's goals for implementation as discussed in Chapter 2. An agency may have a goal to use BMD specifications only on certain types of projects (e.g., new construction, major rehabilitation, mill and overlay) or on projects of a certain size (e.g., tons of asphalt mix > 20,000 tons). Also, an agency may choose to initially use the performance tests only for mix design approval for the first year or so, then work toward adding the performance tests to the QA program in year two or three.

Important factors that should be considered in test selection include field validation, sample preparation, specimen conditioning and testing equipment costs, precision (including both repeatability and reproducibility), material sensitivity, training needs, and overall turn-around time from sampling to results.

For example, the top three factors considered by the New Jersey Department of Transportation (NJDOT) in selecting performance tests for mixture design were field validation, repeatability, and specimen conditioning and testing time. At its current stage of implementation, NJDOT is particularly interested in effective and practical test methods for routine use during production with results that can be tied to a pay factor for acceptance of asphalt mixtures. Both NJDOT and the Texas Department of Transportation (TxDOT) are exploring the use of surrogate tests for acceptance during production with correlation to performance tests used in asphalt mix design (e.g., correlating IDEAL-CT to OT for cracking evaluation).

An agency may elect to conduct specific studies to establish new or modify existing performance tests. This includes the development of a new standard test method or modifying an existing test method. For example, the Illinois DOT supported the research and other activities associated with the development of the Illinois Flexibility Index test (I-FIT) to address observed brittleness of high RAP/RAS mixtures in Illinois. The Louisiana Transpiration Research Center conducted research to develop the Semi-Circular Bend (SCB) Test to discriminate the cracking resistance of polymer modified versus unmodified mixtures in Louisiana. On the other hand, NJDOT supported the research and other activities associated with the use and modification (as needed) of the existing Overlay Test (OT) and Flexural Beam Fatigue (FBF) in specialty asphalt mixtures to address related pavement distresses.

A general tendency is for many highway agencies to add new requirements to their existing mix design and acceptance testing requirements without eliminating unnecessary criteria. That is not the best approach to making significant improvements in asphalt pavement performance and would eventually require much more time and technical manpower to do legacy testing and performance testing on each new mix design. Therefore, part of the effort of the BMD Implementation Committee must also be to reach a consensus on which legacy criteria (e.g., volumetric properties, N_{design}, RAP limits, and others) will be relaxed or eliminated when the mixture performance test criteria are added to the mix design and acceptance specifications. For example, several states have found it necessary to allow a much lower air void limit to increase asphalt contents and improve cracking test results. These changes have had a negligible impact on rutting resistance as evident by both laboratory tests as well as on full-scale accelerated loading on the NCAT Test Track. Another state in the process of adopting the Hamburg Wheel Tracking Test as part of their BMD implementation effort plans to delete the requirements for TSR testing (AASHTO T 283) since the Hamburg test can identify moisture sensitive mixtures.

3.3 Dealing with Reheating and Conditioning/Aging of Mixtures

As part of the plan to implement mixture performance tests, decisions will have to be made on how to deal with mixture conditioning as part of sample preparation. Mixture performance tests are much more sensitive to mixture conditioning than volumetric properties. Even reheating a plant-produced mixture the day after it was sampled can significantly affect most cracking test results, whereas, in most cases, reheating mixtures has a negligible effect on the volumetric results. Therefore, highway agencies and their industry stakeholders must decide how to handle mix conditioning as part of the BMD implementation plan and establish specific policies and standardized procedures for mixture reheating and mix conditioning (i.e. aging) in order to achieve comparable results.

Currently, there is not a national standard for reheating mixtures to compaction temperature and the existing procedures for short-term and long-term aging in AASHTO R 30 are in a state of flux.

A short-term aging procedure is necessary for all lab prepared mixtures to simulate, as much as possible, the aging that occurs during plant mix production, storage (i.e. in a silo), transportation, and placement until it cools to ambient temperatures. Currently, for the preparation of samples for "mechanical property testing", AASHTO R 30-02 (2019) recommends short-term aging that consists of aging loose mix samples for four hours at 135°C. However, based on the findings of NCHRP 9-52 (Newcomb et al., 2015), AASHTO COMP is currently considering reducing short-term aging to two hours at 135°C for hot mix asphalt or at 116°C for warm mix asphalt. If adopted, the change from four hours to two hours would be applicable to the preparation of specimens for rutting tests, cracking tests, and moisture sensitivity tests. The impact of this change on most mixtures will be a decrease in rutting resistance, an increase in cracking resistance, and likely an increase in moisture damage resistance. Agencies that have used the existing short-term aging procedure for mechanical property tests and set their performance test criteria accordingly based on historical lab and field data will have to decide if they will follow the new short-term aging protocol and revise their mix criteria or continue to use the existing short-term aging protocol and keep their established criteria. It is worth noting that several studies that that have compared results from lab prepared mixtures using the existing R 30 short-term aging protocol to test results on the same mixtures produced through plants have shown that there is not a consistent trend to that the plant mixtures are either more or less aged than corresponding lab mixture prepared using R 30 (Chen et al., 2019; Illinois DOT, 2018, Ling and Buchanan, 2022).

It is widely recognized that aging of asphalt pavements continues slowly over time causing the surface layer to become stiffer and more susceptible to cracking and raveling. Numerous studies (Mirza and Witczak, 1995; Koohi et al., 2012; Yin et al., 2017) have demonstrated that the amount of aging decreases exponentially with depth. Therefore, a laboratory protocol to simulate long-term aging is largely considered to be necessary only for surface mixtures. Most stakeholders acknowledge that a long-term mix aging protocol should be used with mix cracking tests that are intended to address surface-initiated cracking to simulate several years of aging in the field. However, several studies (Newcomb et al., 2015; Islam et al., 2015; Howard and Doyle, 2015) have shown that the current long-term aging protocol (compacted specimen aging at 85°C for five days) in AASHTO R 30-02 (2019) does not adequately simulate in-situ aging of surface layers. There is currently a lack of consensus regarding an alternate procedure to simulate in-situ aging and the number of years of service life that the protocol should represent. Most asphalt researchers now seem to prefer loose mix aging rather than compacted specimen aging, but a few agencies still use a compacted specimen aging protocol. Loose mix aging achieves more uniform and accelerated aging of the mixture, whereas compacted specimen aging results in aging from the outside-in; that is - the specimen is baked like a loaf of bread with a crust on the outside. Furthermore, loose mix aging can accommodate elevated aging temperatures without causing specimen distortion issues.

A basic question that researchers and practitioners should consider in establishing an aging protocol is how many years of in-service aging the protocol is trying to simulate. Several national studies (Shen et al., 2017; Bonaquist et al., 2021) have shown that cracking of surface layers typically first begin to appear after about four years in-service. For some asphalt pavements, it may be much later. To account for the differences in aging rates for pavements in hotter and cooler climates, some studies (Yin et al., 2014; Newcomb et al., 2015; Chen et al., 2018) have based the target field aging condition on cumulative degree days (CDD), which is defined as the accumulation of daily high temperatures throughout the layer's service life. An analysis of over 80 pavements showed that top-down cracking began to appear after 70,000 CDD which typically occurs in four to five years in southern states and seven to eight years in many of the northern states bordering Canada (Chen et al., 2018). This demonstrates that the critical time horizon to differentiate cracking susceptibility of surface layers is four to eight years and thus, this time horizon should be targeted in long-term aging protocols.

For any laboratory aging protocol, there is a trade-off between time and temperature. Using a higher aging temperature shortens the amount of time required to achieve a specific aging condition. Earlier studies indicated that aging temperatures above 100°C altered the binder chemistry (oxidation kinetics) in a way that was inconsistent with ambient field aging (Peterson and Harnsberger, 1998; Peterson, 2009; Rad et al., 2017; Kim et al., 2018), although it is not clear if that phenomenon is applicable to all or certain types of asphalt binders. Most mixture aging experiments have used temperatures between 85°C and 135°C with times ranging between 6 hours and 120 hours (5 days). Advantages of using 135°C aging are that it allows labs to use the same oven setting for short-term aging of hot mix asphalt and that the accelerated aging yields a faster overall turnaround of test results, which is desired for mix design and production acceptance purposes. In addition to the potential issue with inconsistent oxidation kinetics noted above, another concern with the 135°C aging temperature is the degradation of styrene-butadiene-styrene block copolymers at elevated temperatures, which will affect the performance properties of the polymer modified binders. Recently, several long-term or mid-term loose-mix aging protocols have been proposed, but none have been widely adopted into practice. This area of research continues to develop.

Including a long-term aging protocol in conjunction with cracking tests for mix design approval of surface mixtures is a prudent step to ensure satisfactory long-term performance of asphalt pavements. Since asphalt binders of the same grade but from different crude sources or those containing different additives can age at substantially different rates, it is not possible to establish a universal relationship between cracking test indices from short-term aged mixtures and long-term aged mixtures. With the potential increased use of recycling agents or rejuvenators for mixtures containing RAP and/or RAS, there is an additional motivation to assess mixture cracking resistance after long-term aging.

For agencies that plan to use performance tests as part of their QA requirements, consideration must be given to two issues: reheating and long-term aging of plant mix samples. As noted above, the effect of reheating on volumetric properties is negligible in most cases, but it can significantly affect the performance test results.

Reheating plant mix samples will be necessary in many cases in order to make performance tests specimens for Quality Assurance. It is essential that agencies establish detailed and standardized reheating procedures so that test results on the same mixture are comparable from lab to lab. If contractors opt to conduct some tests on hot-compacted specimens (i.e., specimens prepared without allowing the mix to drop below the lab compaction temperature after sampling) to expediently assess mix quality, they should consider establishing a "reheating adjustment factor" (RAF) for the test because the hot-compacted specimens are likely to have better cracking but slightly reduced rutting test results than the reheated specimens due to reduced asphalt aging. For example, the RAF could be a ratio of the cracking test result for specimens compacted after reheating the mixture divided by the result of the same mixture using hot compacted specimens. The RAF would then be multiplied by future test results for hot-compacted specimens to estimate the results of reheated specimens. RAFs could be established for rutting and cracking tests if both tests are conducted on hot-compacted specimens. At this time, it is recommended that RAFs be established for each mix design using plant mix samples because the RAF is highly dependent on the aging susceptibility of the mixture.

Similarly, an "aging adjustment factor" (AAF) could be established for surface mixtures to relate cracking test results for reheated specimens and long-term aged specimens. As with the RAF, the AAF should be based on plant mix samples because the binder used during mix design may be different than the binder used during production and that the short-term aging procedure used in mix design may not accurately simulate the aging during plant mix production. Ideally, the AAF would be established during the initial production lot for the mix in the same way that density correction factors are established between cores and density gauges as part of the test strip for the mix design. Using an AAF will help minimize the turnaround time of QA tests during production for agencies that opt to evaluate mixture cracking resistance after long-term aging.

3.4 Performance Test Relationship Confirmation and Criteria Development

A critical step in selecting appropriate mixture performance tests is to make sure that the test results have a strong relationship to field performance, thus supporting the development of specification criteria. The first step is to review and assess the validity and applicability of past studies that related test results to field performance. Very few studies have actually conducted well-controlled, field experiments to establish relationships between test results and field performance. Some BMD test development research has been based on field sites that did not adequately characterize the underlying pavement structures, which is essential to understanding the type of distresses manifesting at the surface. Others compared pavements of different ages or subject to very different traffic loads, which confounded the evaluation of field performance correlation.

To establish laboratory-to-field relationships, it is recommended that a group of field test sections be constructed in which the only variable among the sections is mixture properties for a single layer. All other variables such as underlying support conditions, traffic, layer thicknesses, and age should be as consistent as possible among the test sections so that field performance differences are not confounded by those factors. Each group of test sections needs to be built to target a specific type of distress (e.g., reflection cracking, thermal cracking, moisture damage, or rutting). The asphalt mixtures used in the test sections can be tested using all of the candidate methods being considered by the agency.

An agency may conduct their own lab-to-field correlation efforts by building test sections in their jurisdiction, or they may work with other states in their region to help establish criteria for their local materials, climate, and traffic conditions. This kind of field validation is a significant effort and generally takes several years before long-term field performance data becomes available and is meaningful. So, beginning the planning process early is important.

A key decision in the field validation experimental plan is the number of test sections. A validation experiment with more test sections will yield more data points to establish the correlations between lab performance test results and field performance, which could improve confidence in the test criteria. On the other hand, more test sections will increase the cost of the experiment by requiring more mix design work up front, more testing of the mixtures produced and constructed in each test section, and more time in monitoring pavement responses. Since it is important to include test sections that will have good and poor field performance by including materials that have performance test results both above and below the proposed criteria, a field experiment with six test sections is likely to provide a good balance between cost and robustness. Table 3.1 shows a general matrix of six test sections with asphalt mixtures covering a wide range of rutting and cracking resistance. For example, Section 1 would contain a surface mix with low rutting resistance and low cracking resistance, while Section 6 uses a surface mix with high rutting resistance and cracking resistance. NAPA's online BMD Resource Guide provides guidance on mix design strategies to improve performance test results. With six test sections, it is possible to establish robust lab-to-field correlations for both rutting and cracking performance, as illustrated in Figure 3.2. It is important that the field experiment include mixtures that have a wide range of rutting and cracking field performance so that the data are useful in setting criteria for the lab results that distinguishes good and poor performance.

Table 3.1. Example field validation experimental matrix with six test sections.

Rutting Resistance	Cracking Resistance		
	Low	Medium	High
Low	①	②	
Medium	③		④
High		⑤	⑥

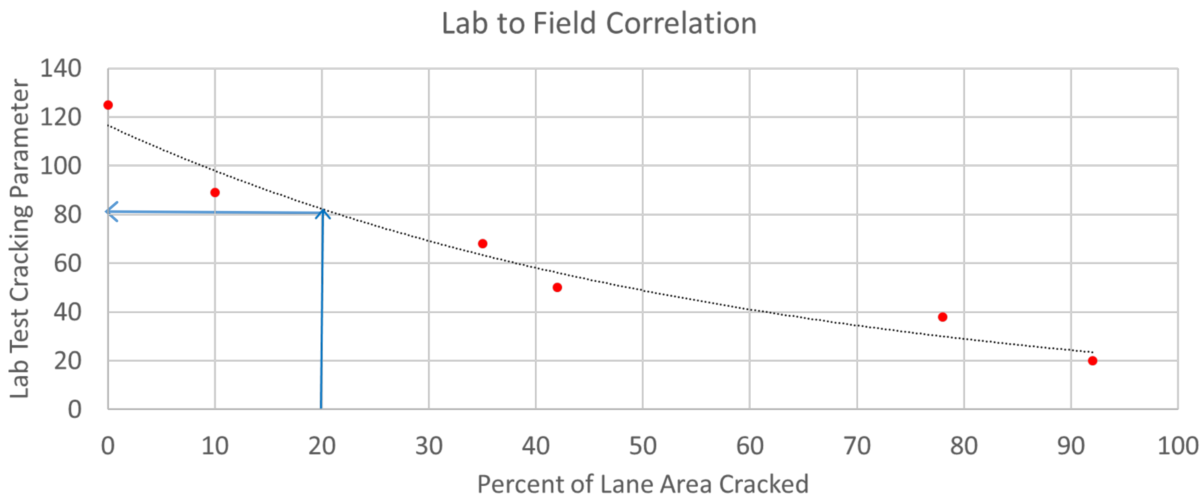


Figure 3.2. Hypothetical results from a validation experiment showing a correlation between a cracking test parameter and field cracking measured in test sections and setting a performance test criteria based on long-term performance expectations.

The agency will have to assume the risk of some test sections performing poorly in order to establish rational criteria for the performance tests. If all of the test sections perform well, then it is not possible to determine where the criteria should be set. The plan for the experiment must include a contingency for repairing or removing and replacing test sections that reach a pre-defined distress condition.

Building the test sections on a single project avoids the performance confounding effects of traffic, aging, and climate conditions. The location of the test sections should be selected so that they will be subject to a consistent traffic speed. The test sections should exclude intersections, have grades below 2%, and should have a consistent number of lanes.

To a significant degree, a project that involves construction of the entire pavement cross-section will provide a consistent pavement structure for the test sections. However, even within a project, varying subgrade conditions can influence load-related distresses. Therefore, it is desirable to choose a project site that has a reasonably consistent subgrade support. When applicable, a pre-construction survey of the project based on non-destructive pavement testing should be conducted to check the consistency of the underlying pavement structure and condition. Suitable field validation experiments can be built as overlays on existing pavements if the condition of the underlying pavement is reasonably uniform throughout the test sections. Mapping of cracks and/or joints in the underlying pavement is necessary to be able to identify reflection cracking that will occur in the test sections. If the overlay is to be constructed on a milled surface, crack mapping should be conducted on the milled surface.

Currently, the Federal Highway Administration (FHWA) has an active pooled fund project to support state highway agencies with the implementation of innovative pavement technologies, products, and processes. The pooled fund is TPF-5(478) *Demonstration to Advance New Pavement Technologies*. One of the topics for consideration is the development of balanced mix design for asphalt pavements. This pooled fund provides a great opportunity for state highway agencies to conduct field validation experiments for BMD performance test criteria. More information about this pooled fund can be found on the Transportation Pooled Fund Program website.

Another decision in the design of the field experiment is the length of the test sections. A commonly used test section length is 500 feet, as used on most Long-Term Pavement Performance (LTPP) experiments and MnROAD cells, although 200-foot test sections have been successfully utilized on the NCAT Test Track. From each section, the first and last 25 feet are transition zones that are excluded from the performance evaluation and are mostly used for extraction of core samples as needed.

Accelerated loading facilities (ALFs), Heavy Vehicle Simulators, and other similar facilities can be used for some field validations, but there are significant limitations with these experiments. The wheel speeds of these facilities are much slower than typical highway traffic and since asphalt is a visco-elastic material, response to loads is highly dependent on rate of loading. Therefore, the results of the test sections from these loading facilities may provide a useful ranking of mix performance, but they are not suitable for setting performance test criteria. Other issues with these facilities are temperature control and aging. Typically, experiments using these loading facilities are conducted at a specific temperature and therefore do not simulate the effects of daily and seasonal temperature changes on pavement performance. Aging of the test sections will occur over time, so an experiment that loads one test section at a time over a period of a few years will be confounded by different degrees of aging in the surface mix.

Full-scale pavement test facilities, such as MnROAD and the NCAT Test Track are well suited for field validation experiments. Test sections at these facilities can be built with consistent underlying support, mix designs can be tailored to meet specific performance test targets, construction is very well documented, traffic loading is known, and performance of the test sections is closely monitored using the same methods used by agencies for assessing the condition of their highway network. Such experiments can be collectively funded by several states to share the cost of the work.

Mixture testing for the validation experiment should include, at a minimum, tests on lab-produced mixtures (representing mix design) and plant-produced mixtures including traditional mixture properties and the selected performance tests. If the Technical Committee is still considering several performance tests or surrogate tests for mix design screening or QA testing, then those tests should be included in the testing plan. Lastly, if the field validation experiment is intended to establish relationships between cracking of surface mixtures and lab tests, then the experimental plan should include testing the mixtures at multiple aging conditions. The laboratory work plan will likely be extensive thus requiring a commitment of significant resources by the agency, the contractor involved in the project, and possibly a third party such as a research organization.

References

Bonaquist, R., J.J. Adams, and D.A. Anderson (2021) Asphalt Binder Aging Methods to Accurately Reflect Mixture Aging, National Cooperative Highway Research Program, Project 09-67.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, 2672(28), pp.359-371.

Chen, C., Moore, N., West, R., Yin, F., & Taylor, A. (2019). Evaluation and comparison of cracking tests for top-down cracking resistance of asphalt mixtures. Journal of the Association of Asphalt Paving Technologists.

Hajj, E.Y., A.J.T. Hand, R. Chkaiban, and T.B. Aschenbrener (2019) Index-Based Tests for Performance Engineered Mixture Designs for Asphalt Pavements, Report No. FHWA-HIF-19-103.Federal Highway Administration.

Howard, I. L., and J. D. Doyle (2015) Durability Indexes via Cantabro Testing for Unaged, Laboratory-Conditioned, and One-Year Outdoor-Aged Asphalt Concrete. Presented at 94th Annual Meeting of the Transportation Research Board.

Illinois DOT Balanced Mix Design Implementation Effort (2018), http://www.pavetrack.com/180925/2018.09%20-%20Day%202%20-%20Unit%207%20-%20Illinois%20BMD%20Implementation.pdf

Islam, M. R., M. I. Hossain, and R. A. Tarefder (2015) A Study of Asphalt Aging Using Indirect Tensile Strength Test. Construction and Building Materials, Vol. 95, pp. 218–223.

Kim, Y.R., C. Castorena, M.D. Elwardany, F.Y. Rad, S. Underwood, G. Akshay, P. Gudipudi, M.J. Farrar, and R.R. Glaser (2018) Long-term aging of asphalt mixtures for performance testing and prediction.

Koohi, Y., J. J. Lawrence, R. Luo, and R.L. Lytton (2012) Complex Stiffness Gradient Estimation of Field-Aged Asphalt Concrete Layers Using the Direct Tension Test. Journal of Materials in Civil Engineering, Vol. 24, No. 7, pp. 832–841.

Ling, C. and S. Buchanan (2022) Implementation considerations of Balanced Mix Design in practice: recent experience in Vermont, Road Materials and Pavement Design, DOI:10.1080/14680629.2022.2029548.

Marasteanu, M., W. Buttlar, H. Bahia, and C. Williams, et al. (2012) Investigation of Low Temperature Cracking in Asphalt Mixtures, National Pooled-Fund Study – Phase II, Minnesota Department of Transportation, Research Services.

Mirza, M.W., and M.W. Witczak (1995) Development of a Global Aging System for Short and Long Term Aging of Asphalt Cements. Journal of the Association of Asphalt Paving Technologists, Vol. 64, pp. 393–430.

Newcomb, D., A.E. Martin, F. Yin, E. Arambula, E.S. Park, A. Chowdhury, R. Brown, C. Rodezno, N. Tran, E. Coleri, and D. Jones (2015) *Short-term laboratory conditioning of asphalt mixtures,* National Cooperative Highway Research Program, Project 09-52.

Petersen, J.C., and P.M. Harnsberger (1998) Asphalt aging: dual oxidation mechanism and its interrelationships with asphalt composition and oxidative age hardening. Transportation Research Record, 1638(1), pp.47-55.

Petersen, J.C. (2009) A review of the fundamentals of asphalt oxidation: chemical, physicochemical, physical property, and durability relationships. Transportation Research Circular, (E-C140).

Rad, F.Y., M.D. Elwardany, C. Castorena, and Y.R. Kim (2017) Investigation of proper long-term laboratory aging temperature for performance testing of asphalt concrete. Construction and Building Materials, 147, pp.616-629.

Shen, S., S. Wu, W. Zhang, L. Mohammad, and B. Muhunthan (2017) Long-term field performance of warm mix asphalt technologies. National Cooperative Highway Research Program, Project 9-47A.

West, R., C. Rodezno, F. Leiva, and F. Yin (2018) Development of a Framework for Balanced Mix Design. NCHRP Project 20-07/Task 406 Final Report. 167p. National Cooperative Highway Research Program, Transportation Research Board.

West, R., D. Timm, B. Powell, N. Tran, F. Yin, B. Bowers, C. Rodezno, F. Leiva, A. Vargas, F. Gu, R. Moraes, and M. Nakhaei (2021) Phase VII (2018-2021) NCAT Test Track Findings, NCAT Report 21-03.

Yin, F., A.E. Martin, E. Arámbula-Mercado, and D. Newcomb (2017) Characterization of non-uniform field aging in asphalt pavements. Construction and Building Materials, 153, pp.607-615.

Yin, F., E. Arámbula-Mercado, A. Epps Martin, D. Newcomb, and N. Tran (2017) Long-term ageing of asphalt mixtures. Road Materials and Pavement Design, 18(sup1), pp.2-27.

Yin, F., and R. West (2021) Balanced Mix Design Resource Guide. NAPA Information Series 143; online version available at https://www.asphaltpavement.org/expertise/engineering/resources/bmd-resource-guide.

Task 3 Checklist of steps for selecting performance tests and establishing performance criteria

- ☐ Identify asphalt pavement distresses that performance tests should address
  - Analyze the agency's Pavement Management System data
  - Conduct network level summary of pavement issues from the project level coring program
  - Survey the agency's regional engineers to collect information on causes of pavement distresses in their jurisdictions.
- ☐ Identify candidate performance tests
  - Use existing method(s)
  - Modify existing method(s)
  - Develop new test method(s)
- ☐ Prioritize factors to be used in selecting the performance tests
- ☐ Select the performance tests that are expected to indicate the resistance of asphalt mixtures to the distresses of concern
- ☐ Decide which legacy criteria (e.g. volumetric properties, N_{design} levels, RAP limits, film thickness, etc.) can be eliminated when the performance test criteria are added to the specifications.
- ☐ Decide how the effect aging will be included in the assessment of cracking resistance of surface mixtures
- ☐ Review literature for lab to field relationships or correlations
  - Determine if the existing information is suitable for setting performance criteria for the state or local highway agency. If not proceed with the remainder of the checklist.
- ☐ Develop an experimental plan for field validation of performance criteria
  - Consider applying for TPF-5(478) for support to build and evaluate a validation experiment
  - Consider working with other regional agencies to share the cost and expertise needed for conducting field validation experiments
  - Construct field validation experiment(s); analyze results and set preliminary performance criteria. This is likely to take several years for the performance differences of the test sections to be meaningful enough to establish useful correlations.

Chapter 4. Performance Testing Equipment and Staffing: Acquiring Equipment, Managing Resources, Training and Evaluating

Objective: To explain the different factors that can be considered in the acquisition of equipment, staffing, and initial training of personnel.

Resources:

- Existing asphalt technician training programs that can be augmented for initial training for the new performance tests.
- Training offered by third party organizations and equipment manufacturers.
- National or regional round-robin (interlaboratory studies) to assess lab proficiencies with the new tests.
- LASTRDA Partners Hamburg Wheel Tracking Machine Comparison

Outcomes: (1) Development of plans by stakeholder organizations for the acquisition of equipment and the utilization of available space and workforce to implement the performance tests and specifications. (2) Plan and execute initial training of personnel who will be involved in interlaboratory studies and establishing baseline data for the development of specification criteria.

4.1 Acquiring Equipment

Equipment needed to conduct the selected performance tests may include new test equipment or modifications to existing equipment as well as equipment for sample preparation, aging/conditioning, and specimen fabrication (e.g., table saws, conditioning chambers, core drills, water baths, aging ovens, etc.). Each lab conducting the tests will need to consider whether existing equipment can be used, modified, or retrofitted, or if new equipment will have to be purchased.

When equipment is acquired within a state that comes from different manufacturers for the same test, an experiment should be conducted with units from each manufacturer to verify that results from each unit are similar. For example, NCAT completed a study comparing six popular pieces of equipment for conducting the Indirect Tension Cracking Test (ASTM D8225) and found that five of the six devices yielded results that were practically equivalent (Moore et al., 2021). Although this kind of variability information may be identified by interlaboratory studies discussed in Subtask 4.5, it is better to know if a particular brand or model of equipment biases results or increases variability before significant purchases are made.

LASTRADA Partners provides a useful comparison of the top manufacturers of Hamburg Wheel Tracking machines to provide potential buyers information to make an informed purchasing decision. The Hamburg manufacturers were asked the same eight questions to provide information on each respective company, the features and safety measures of their Hamburg equipment, compliance with current standards, calibration and maintenance recommendations, service and support options, and requirements for installation. The website link provided above under "Resources" provides video interviews with each company's representative and a brief statement about the company and their Hamburg equipment.

Mohammad et al (2017) also provided a comprehensive evaluation of Hamburg machines to assess their compliance with the requirements of AASHTO T 324.

When several laboratories need to purchase the equipment, there may be cost savings with purchasing a large number of units at the same time, such as with a pooled-fund equipment purchase. Pooled-fund purchases and group purchases were common during the early stages of Superpave implementation.

4.2 Managing Resources

Prior to installing any testing or sample preparation equipment, an organization will need to determine the potential impact that performance testing will have on their existing staffing capabilities. Additional testing may require additional laboratory staff or a reallocation of duties to existing staff. Some agencies have accomplished this through consultant staffing augmentation (either in-house or externally), or through the elimination of less-critical duties for staff.

With respect to equipment acquisition, an organization will need to determine whether their existing labs will require any type of rearrangement/remodeling or changes to their electrical, plumbing or ventilation systems. For example, some Illinois DOT District laboratories rearranged their workspaces to improve laboratory efficiency. Another state DOT converted stairwell space into a room for new equipment and converted a closet into a space for coring and sawing specimens.

In some cases, space and labor limitations may dictate how BMD specifications are implemented if the available resources only allow a limited amount of performance testing to support on-going construction projects. The transition to using the new performance tests needs to be accomplished while continuing to meet the testing needs of regular projects.

4.3 Conducting Initial Training

As new performance tests are adopted, technicians conducting the tests need to understand the test equipment, sample conditioning protocols, specimen preparation procedures, test methods, data analysis, and interpretation. If plant mix samples are to be used in the BMD specification, training also needs to include protocols for handling samples between the point of sampling and the testing laboratory. Initially, informal training would need to be provided and would likely include information from several sources including written procedures, demonstrations from equipment manufacturers, BMD workshops, and discussions with other users of the tests. Third-party training courses will also be helpful, especially those containing hands-on lab sessions. As senior laboratory technicians, laboratory managers, and supervisors are trained and become more experienced in running the tests, they will likely be involved in future staff training. The initial training can be limited just to the personnel and laboratories participating in inter-laboratory studies (Subtask 4.5), benchmarking studies (Subtask 5.2), and shadow projects (Subtask 5.3). Training of the state's entire asphalt QA workforce will be discussed in Chapter 7.

4.4 Refining Procedures

Chapter 3 identified several factors to be considered in selecting the appropriate performance tests. Other considerations deal with how the performance tests can be put into routine practice. Research gaps may need to be identified and work conducted in order to complete missing and/or incomplete information, or existing procedures and methodologies may need to be refined. Examples of additional considerations for refinement include:

- Refining the performance test methods with clear and specific details on specimen conditioning, preparation, calculations and data analysis methods, standardized reporting of test results and parameters, as well as database attributes for the stored raw/primary test data. For example, differences may exist in data collection by different equipment manufacturers. In some cases, a standardized program or worksheet is needed to ensure that results are calculated/analyzed in a consistent manner. An example of this issue is the collection of deformation data from different manufacturers of the Hamburg Wheel Tracking Device. Furthermore, AASHTO T 324 does not provide guidance on selecting the test temperature or guidance on how to determine the steady-state portion of the deformation versus curve(s), which is critical to determining the stripping inflection point (SIP).
- Establishing consistent sample conditioning, handling, fabrication, and preparation procedures to reduce lab-to-lab differences and overall variability of test results.
- Understanding the differences between laboratory and plant-produced asphalt mixtures is needed to understand the changes that occur to an asphalt mixture between mix design and production, and why performance tests must also be included in QA testing and how it can be done in an effective and practical manner.
- Establishing laboratory aging and conditioning protocols for asphalt mixtures prior to performance testing. This involves the selection of short- and long-term aging protocols for the performance tests. The aging protocol may differ for laboratory- and plant-produced asphalt mixtures as well as surface mixtures and mixtures used in underlying layers.

4.5 Conducting Interlaboratory Studies

One of the factors in selecting the appropriate performance test (Subtask 3.2) is test precision. A test with poor precision may be unable to discern good mixtures from bad mixtures. For tests whose results are used for materials approval and/or acceptance, it is necessary to establish the method's precision and bias information. The proper method for determining test precision information is through an inter-laboratory study, also known as a round-robin study, as described in ASTM E691. State DOT and industry laboratories should participate in the inter-laboratory studies after becoming familiar with the tests.

Interlaboratory studies are helpful in two ways. First, the participating labs get useful insight as to how their results compare with other labs, which can help identify if errors are being made in preparing the samples and conducting the test. Second, the combined results from all the labs can be used to establish the acceptable range of two test results from a single operator (i.e., within-lab) and the acceptable range of split-sample results from two different laboratories (i.e., between-lab). Knowing the within-lab and between-lab test variabilities of different candidate tests determined using ASTM E691 is useful information to help select the most preferred test option and can also be used in the SHA's Independent Assurance program. Agencies that utilize the same test procedures can combine efforts on interlaboratory studies and/or share the results.

Interlaboratory studies may be conducted nationally (AASHTO resource, academia, etc.) or locally at the state level. An interlaboratory study is typically designed to include a minimum number of laboratories, number of samples and replicates, sample preparation, and the various tests that are conducted on the samples. As an example, recently the Virginia Department of Transportation (VDOT) conducted an interlaboratory study for the IDEAL-CT test. The study included 50 laboratories and two different asphalt mixtures. The mixtures were fabricated all at one laboratory, and uncompacted samples were then sent to each of the laboratories involved in the study, where they were conditioned, compacted, and tested. The results were then returned to VDOT for analysis. In 2018 and 2019, NCAT conducted a preliminary round-robin study for several popular rutting and cracking tests being considered for BMD. One key finding from that study was that a major source of variability can be attributed to specimen preparation.

For most of the performance tests being considered for BMD implementation, some information on the repeatability (within-laboratory variability) is available in the test summaries in the online BMD Resource Guide, however, it is somewhat limited. Although this information is commonly used by organizations as an assessment of variability in the selection of tests as discussed in Chapter 3, the within-lab variability is only part of the variability information needed for implementation.

Another part of assessing a test procedure's variability is a ruggedness study, conducted in accordance with ASTM E1169. The purpose of a ruggedness study is to identify those factors that strongly influence the measurements provided by a specific test method and to estimate how closely those factors need to be controlled. For example, if a test method requires a test to be performed at 7.0% air voids, a ruggedness study will determine what the allowable tolerance should be on air voids. Ruggedness studies should be conducted prior to the finalization of the test method and before any interlaboratory studies are conducted and are typically conducted nationally. Unfortunately, many of the tests that are currently in use for materials approval and acceptance have not gone through a formal ruggedness experiment. The recently completed NCHRP Report 987, *Ruggedness of Laboratory Tests to Assess Cracking Resistance of Asphalt Mixtures* provides details of the ruggedness experiments as well as recommendations to improve the following eight cracking tests:

- Illinois Flexibility Index Test (I-FIT), AASHTO TP 124
- IDEAL-Cracking Test, ASTM D8225
- Overlay Test, Tex-248-F
- Disc-Shaped Compact Tension Test, ASTM D7313-13
- Fracture Energy Using the Semicircular Bend Geometry (Low Temperature), AASHTO TP 105
- Semi-Circular Bend Test at Intermediate Temperatures (Louisiana method) , ASTM D8044
- University of Florida Indirect Tensile Test
  Bending Beam Fatigue, AASHTO T321

A ruggedness study was also recently completed the cyclic fatigue test (AASHTO TP 107-18 and TP 133-19) using an Asphalt Mixture Performance Tester (AMPT) (Castorena et al. 2021).

References

ASTM E691, Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method, ASTM International.

ASTM E1169, Standard Practice for Conducting Ruggedness Tests, ASTM International.

Castorena, C, B.S. Underwood, Y.R. Kim, K. Lee, N. Tran, and A. Taylor (2021) Ruggedness and Interlaboratory Studies for Asphalt Mixture Performance Tester (AMPT) Cyclic Fatigue Test: Phase I Report, FHWA-HRT-21-057.

LASTRDA Partners Hamburg Wheel Tracking Machine Comparison: https://info.lastradapartners.com/videos/hamburg-machine-comparison

Mohammad, L., M. Elseifi, W. Cao, A. Raghavendra, and M. Ye, (2017) Evaluation of various Hamburg wheel-tracking devices and AASHTO T 324 specification for rutting testing of asphalt mixtures, Road Materials and Pavement Design, Vol. 18 – Issue sup4, 92nd Association of Asphalt Paving Technologists Annual Meeting.

Moore, N., R. Steger, B. Bowers, and A. Taylor (2021) Investigation of IDEAL-CT Device Equivalence: Are All Devices Equal? accepted for publication, Journal of the Association of Asphalt Paving Technologists.

Taylor, A., Moore, J., and Moore, N., (2021) NCAT Performance Testing Round Robin, NCAT Report 22-01. National Center for Asphalt Technology, Auburn, AL.

Zhou, F, D. Newcomb, and S. Hu, (2022) NCHRP Report 987, Ruggedness of Laboratory Tests for Asphalt Mixture Cracking Resistance, The National Academies Press. https://doi.org/10.17226/26528.

Chapter 5. Establishing Baseline Data

Objectives:

1) To gather and analyze data using the new mixture performance tests needed to establish appropriate specification criteria, and
2) To considers several sources of variability.

Resources:

- Checklist for Task 5 Establishing Baseline Data, see p. 39

Outcomes: (1) Assessment of potential pavement management data to aid in setting criteria for mixture performance tests. (2) Development of a plan for benchmarking studies. (2) Analyses of benchmarking data. (3) Planning and execution of shadow projects. (4) Analyses of shadow project data to determine within-lot pooled standard deviations for the selected performance tests. (5) Preliminary assessment of the logistics of conducting the performance tests for QA testing.

Establishing baseline data incudes five steps that are essential in the development of performance test criteria and related performance specifications. The baseline data will help both the highway agency and contractors gain confidence that the test criteria used for asphalt mixture design and/or acceptance are appropriately set. Ultimately, the test criteria can be established from information collected and analyzed as described in Subtasks 3.3, 4.5, 5.2, 5.4, and 6.5.

5.1 Reviewing Existing Lab Results & Pavement Management System Data

The first step in establishing baseline data is for the highway agency to leverage its existing information from performance tests on asphalt mixtures in its jurisdiction (if available). This involves reviewing, organizing, and analyzing historical test data for asphalt mixtures and their associated field pavement performance from the agency's pavement management system. Analysis of historical data will allow the highway agency to explore the applicability and/or validity of performance testing for use in a BMD specification and identify needed changes to existing test methods and criteria. Basically, this effort should be used to answer the question "have pavements built using mixtures evaluated with performance tests performed as expected over a sufficient period of time?"

A key component of this step is the development of a database of performance testing results. Having easy access to test data is critical for both short- and long-term implementation. This could include test data from in-house studies, research projects, mix design verification, shadow projects, as well as pilot projects. The SHA should establish early in the implementation process what sample and test data needs to be collected, and also how it will be stored and accessed in the future. Ideally the information would be compiled in the SHA's existing materials database.

5.2 Conducting Benchmarking Studies

The second step in establishing baseline data is conducting benchmarking studies. Benchmarking refers to performance testing a representative set of currently approved asphalt mixtures used in the state or local jurisdiction to understand the range of results for the selected tests. Benchmarking studies can evaluate results from several of the promising performance tests to help determine which test provides the best information and is most implementable. It is desirable for benchmarking to include two sets of data, one set consisting of lab-produced mixture samples such as would be prepared for mix design development and approval, and a second set consisting of plant-produced mixture samples such as those used for QA testing. Although some contractors will likely do their own benchmarking analyses, the analyses discussed in this section focuses on the development of a broad database of mixtures tested in a single laboratory such as the agency's central laboratory or a single third-part laboratory. Using a single laboratory's results for this analysis will avoid lab-to-lab variability issues. Nonetheless, the agency should collaborate with contractors and testing labs to establish the benchmarking plan.

It is desirable to include mixtures with known field performance when selecting mix designs for the benchmarking experiment. For example, NJDOT's Pavement Management System was used to provide useful information on good and poor performing asphalt mixtures. The benchmarking plan needs to include asphalt mixtures from all of the state's mix categories as well as mixtures with different asphalt binder grades, aggregate types, and recycled materials contents. In general, it is desired to build the database to include 15 to 25 mixtures in the most commonly specified categories such as surface mixes used on medium traffic projects. For the analysis to be meaningful, at least five mixes should be included for the lesser used mix categories. In 2020, the Colorado DOT requested a bucket of mix from every project being constructed in the state and began fabrication and testing specimens from those samples during the winter months to build their benchmarking database.

Some states have asked that contractors compact or fabricate the specimens for the benchmarking effort. Because sample fabrication has a significant impact on mixture performance test results, the organization leading the benchmarking experiment should provide a detailed, step-by-step sample fabrication procedure to all organizations that will prepare samples for the benchmarking testing. However, if possible, it is preferred to have all specimens used in the benchmarking effort to be prepared in the same lab.

Once benchmarking testing is completed, the database of performance test results can be analyzed to determine two important pieces of information: (1) the distribution of results for each of the major mix categories (i.e., surface mixes in low, medium, and high traffic categories), and (2) the effects of key mix components (i.e., recycled materials contents, binder grades, aggregate type, etc.) on the performance test results.

Figure 5.1 shows a cumulative distribution function (CDF) for benchmarking results of a cracking test for a state highway agency. This chart is somewhat like a gradation plot of the percentage passing on the y-axis and sieve size on the x-axis, except that here the results show the percentage of CT_{Index} results below any value over the range of CT_{Index} values obtained from the benchmarking study. From this kind of chart, the median value (50th percentile) for the mixtures in the state or the percentage of mixes in the state that would pass or fail any particular CT_{Index} value can easily be determined. For example, the CDF of CT_{index} in Figure 5.1 shows that if the state set a minimum CT_{index} of 50, then approximately half of the current mix designs would not meet that criterion. This kind of chart will be very important to understand the impact of setting a particular specification criterion for the agency. It is not recommended for highway agencies to set the specification criterion by selecting an arbitrary percentile on the CDF curve. Instead, the criterion should be selected based on the correlation of performance test results to field performance data from existing projects, test sections in field validation experiments, and/or accelerated pavement testing facilities. If suitable field performance data are not available, then criteria based on benchmarking studies could be used, however, this criteria should be considered preliminary and will need to be reviewed and adjusted as correlations between the selected performance tests and field performance are developed.

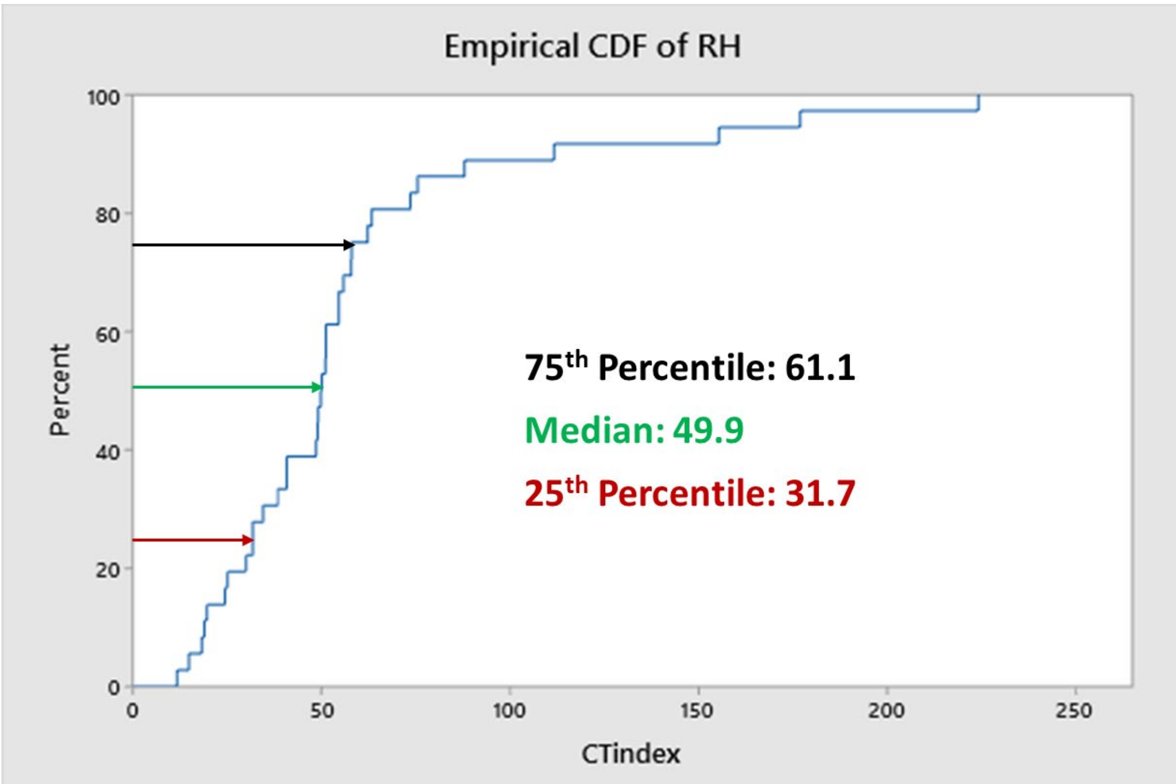


Figure 5.1. Example cumulative distribution function of Reheated (RH) CT_{Index} from a benchmarking study.

CDF plots of each traffic category on the same chart may be useful to understand the impact that existing specifications may have on cracking and rutting resistance. For instance, a state that requires different gyratory compactive efforts for mix designs for various traffic categories may be able to easily identify the impact of design compactive effort on the rutting and cracking resistance of its mixtures.

Another type of useful analysis of benchmarking data is the use of boxplots to assess how key mix components, such as recycled materials contents, binder grades, and aggregate type, affect performance test results. Figure 5.2 shows an example boxplot of CT_{Index} results with the data sorted by the primary aggregate types used in mix designs for a state. In this case, granite mixes had significantly higher CT_{Index} results and thus, were expected to have better cracking resistance than gravel, limestone, and quartz mixes.

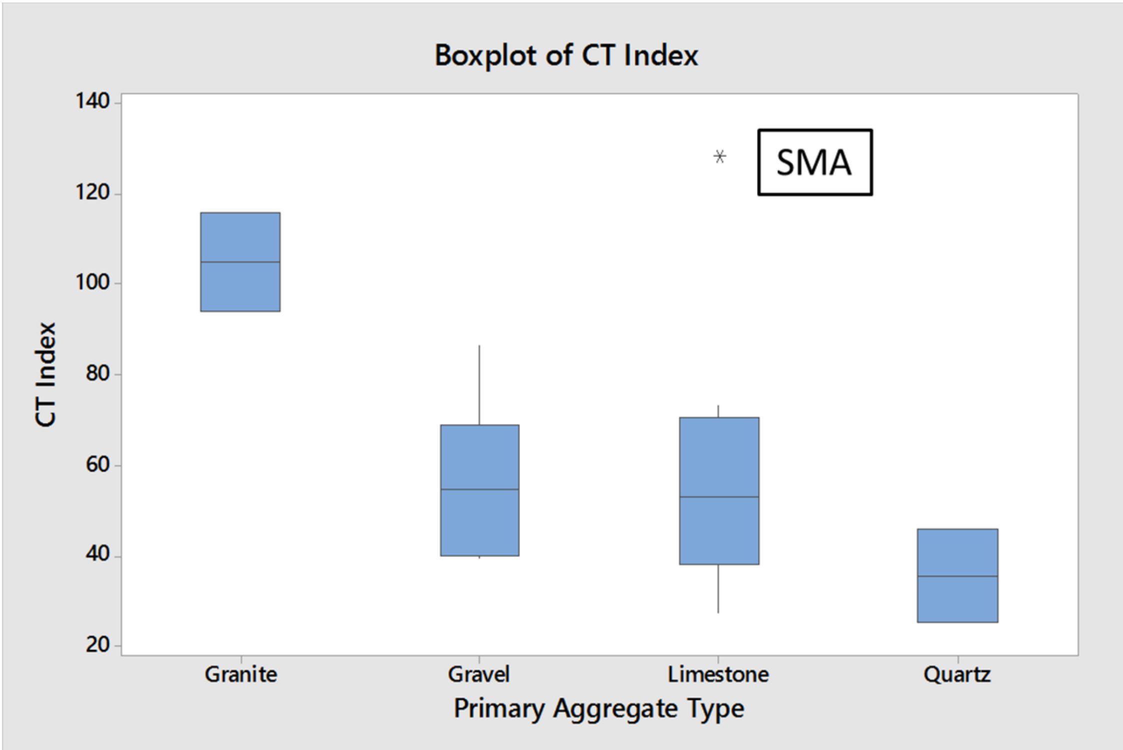


Figure 5.2. Example boxplot of CT_{Index} results sorted by primary aggregate types in mixtures

In general, these types of graphical analyses are well suited for the purposes of the benchmarking studies and communicating the results with stakeholder groups. Statistical analyses, such as Analysis of Variance and multiple comparison tests may also be useful, but it can be more challenging to explain to practitioners who do not have a strong background in statistical analysis.

5.3 Conducting Shadow Projects

The third step in establishing baseline data is conducting shadow projects. Shadow projects are highway paving projects on which the selected performance tests are conducted in the background, so to speak, to gather information needed for developing specification criteria on future projects. In other words, the results of the performance tests are not used to accept or reject materials, adjust payment, or even to motivate mix production changes. The three goals of the shadow projects are: (1) to better familiarize both agency (or agency representative) and contractor personnel with the selected tests, (2) to add to the database of test results from the benchmarking studies, and (3) to gather information on typical production variability.

Although the shadow projects are conducted on projects where the agency's existing specifications are used for acceptance testing, additional QA manpower and testing resources will be needed to conduct the performance testing. Ideally, the performance testing conducted on the shadow projects should be performed by agency personnel for states that only use state test results in acceptance decisions, or by contractor personnel in states that allow the use of contractor data in acceptance decisions. In either case, conducting the performance tests by both the agency and the contractor is a good idea because the results can be compared to assess between-lab variability. Targeted training will be necessary for personnel unfamiliar with the new tests and specimen preparation details including appropriate mixture conditioning procedures.

In general, sampling of mixtures for performance testing should occur at the same time and frequency used for existing acceptance testing. This will allow for a comparison of how performance test results vary along with the traditional acceptance properties as well as provide some evidence as to the causes of the variations in the performance test results. It will also give stakeholders a sense of the logistics necessary to conduct the selected performance tests at a particular frequency for acceptance in the future.

As noted above, one of the goals of shadow projects is to gather information on typical production variability, also known as process variability. AASHTO R 9, *Standard Practice for Acceptance Sampling Plans for Highway Construction*, recommends that quantifying process variability be based on a "large number of project data." This standard provides an example that used 10 projects across a state, and for each project, data from 30 paving days was used to generate "typical standard deviations" for the quality characteristic of interest. R 9 also states that the appropriate variability measure for developing the acceptance specifications is the "within-lot pooled standard deviation". In other words, test results from each project should be used to determine standard deviations based on the proposed lot size. The within-lot standard deviations from each project are then pooled together for a representative value to use in setting specification limits (An example of this can be found in the Appendix of R 9). The rationale for this approach is that within-lot standard deviations will likely differ from project to project depending on the consistency of the component materials, plant operations, and skill levels of the technicians involved in sampling mixtures and conducting the tests. Furthermore, it is desirable to include projects that utilize different mix types as that may affect both the mean values and the standard deviations for the performance test results.

Selection of projects for shadow testing should consider using projects that have enough of the targeted mix type for collection and testing of three to four full lots of mixture. Considering that most states use lots divided into four or five sublots, this would require between 12 and 20 tests for each of the selected performance tests per project. With this scenario, sampling from 10 shadow projects would generate 120 to 200 tests representing 30 to 40 lots from which to calculate within-lot pooled standard deviations. This would be more than sufficient for determining the within-lot pooled standard deviation for the performance tests.

At the completion of each shadow project, the data from this additional testing would be summarized, shared, and discussed with the agency's project personnel and contractor personnel involved in QA.

5.4 Analyzing Production Data

One of the reasons for conducting shadow projects is to collect data on production variability of the performance test results. Since the tests may be new, there may be limited data from other agencies with which to compare. If an agency considers variability data from other states, it is important that the variability calculations (i.e. standard deviations) be based on the same test, sampling reheating and conditioning procedures, and the same number of tests per lot that the agency plans to use.

In order to determine if the variability from each of the shadow projects is due to the new sampling and testing methods or if it is normal production variability, the variability of the traditional acceptance quality characteristics (AQCs) should first be analyzed. Table 5.1 summarizes the range and average standard deviations for volumetric properties and aggregate gradations from typical asphalt mixture production data based on field projects from 11 states (Mohammad et al. 2016). These results agree with the research findings of an ALDOT study (Ajede et al., 2018) and a FLDOT study (Sholar et al., 2003). If any of the traditional AQC shadow project data are outside of these ranges (assuming they are representative of the industry's current practice for asphalt mixture production and construction), that project may not be "in control" and the standard deviations from the new performance test data from that project may be suspect.

Table 5.1. Summary of Standard Deviations for Volumetric Properties and Gradations from Production and Construction Data (Mohammad et al. 2016).

Property	Range of St. Dev.	Avg.
Asphalt Content, %	0.17 – 0.29	0.20
Air Voids, %	0.33 – 0.99	0.62
VMA, %	0.38 – 0.64	0.54
VFA, %	3.40 – 4.92	4.03
G_{mb} (lab compacted)	0.008 – 0.018	0.015
G_{mb} (cores)	0.008 – 0.033	0.019
G_{mm}	0.005 – 0.012	0.011
Field Density (%G_{mm})	0.74 – 1.49	1.11
Percent Passing Sieve	Range of St. Dev.	Avg.
25.0 mm	1.55 – 2.66	1.86
19.0 mm	0.93 – 2.59	1.77
12.5 mm	0.99 – 3.54	2.17
9.5 mm	1.50 – 3.75	2.35
4.75 mm	1.87 – 3.48	2.62
2.36 mm	1.62 – 2.62	2.20
1.18 mm	1.70 – 2.05	1.81
0.60 mm	1.43 – 1.84	1.60
0.30 mm	1.07 – 1.22	1.16
0.15 mm	0.80 – 0.99	0.87
0.075 mm	0.32 – 0.84	0.55

A recommended procedure for evaluating potential outlier results from a set of sublot tests in a lot is provided in Appendix A.

Another important check of the production data is an assessment of normality (i.e., that results follow a normal distribution), since normality is an underlying assumption for most statistical acceptance procedures for construction materials. A visual observation of the distribution of proposed AQCs from the shadow projects is a basic check. Many statistical software programs provide normality tests such as Anderson-Darling, Shapiro-Wilk, and Kolmogorov-Smirnov tests. Both graphical interpretations of distributions and statistical tests of normality are better when more data points are available. An example normality check from an actual dataset of Flexibility Index results from shadow projects is provided in Appendix B.

As discussed in the previous section, one of the primary goals of shadow projects is to determine "typical standard deviations" for the new tests that are intended to be used as AQCs. If the acceptance plan will be based on lot–by–lot acceptance, then the variability that is used to establish the specification limits must be that which is appropriate for a typical lot. Therefore, the correct method for determining "typical standard deviations" is to determine within-lot pooled standard deviations from representative projects across the state. Table 5.2 shows example data from a state that conducted eight shadow projects in 2019 in which they tested traditional AQCs of asphalt content and air voids as well as rutting with the Hamburg Wheel Tracking Test and the Flexibility Index cracking parameter on each project. In this dataset, lots were defined as the quantity of a mix for the entire project.

Table 5.2. Example of Within-Lot Standard Deviations for Asphalt Tests from Shadow Projects in a State.

Project No.	No. of Results	Project Statistics	Asphalt Content, %	Air Voids, %	Hamburg Rut Depth, mm	Flexibility Index
1	8	Avg.	5.84	4.45	3.29	16.14
		Std. Dev.	0.14	0.67	0.36	2.64
2	13	Avg.	6.08	3.95	3.55	10.20
		Std. Dev.	0.18	0.47	1.80	2.64
3	14	Avg.	6.14	4.09	4.40	14.26
		Std. Dev.	0.30	0.51	1.98	3.02
4	9	Avg.	5.30	3.68	4.00	14.64
		Std. Dev.	0.17	0.67	0.56	3.41
5	10	Avg.	4.85	4.08	2.79	6.23
		Std. Dev.	0.15	0.52	0.38	1.40
6	8	Avg.	6.07	3.90	3.07	11.29
		Std. Dev.	0.31	0.70	0.51	1.33
7	6	Avg.	6.17	3.88	3.39	12.95
		Std. Dev.	0.20	0.63	0.40	1.65
8	6	Avg.	4.80	3.73	3.81	6.70
		Std. Dev.	0.06	0.48	0.41	1.71
Pooled Standard Deviation			0.21	0.41	0.88	2.35

The pooled standard deviation is a method for estimating a single standard deviation that represents all of the independent groups of data from the shadow projects. It is a weighted average of each group's (project's) standard deviation. The weighting is based on the number of samples used in each group (project) and gives larger groups a proportionally greater effect on the overall estimate. Equation 5.1 is the simplified way to calculate the pooled standard deviation.

$$s_{pooled} = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2+\cdots+(n_k-1)s_k^2}{n_1+n_2+\cdots+n_k-k}} \qquad \text{Eq. 5.1}$$

Where s_{pooled} is the pooled standard deviation, n_1 is the number of samples/results from project/group 1, s_1 is the standard deviation from project/group 1, and k is the total number of projects/groups.

To help demonstrate the impact of production variability, an example of using the within-lot pooled standard deviation in a percent within limits (PWL) specification is provided in Appendix C.

Further analysis of the variability of an AQC may provide motivation to further explore partitioning the total variability into the three components (testing variability, sampling variability, and materials variability) so that subsequent efforts can appropriately focus on improving the component(s) with the highest impact. This would be a worthwhile future research effort.

5.5 Determining How to Adjust Asphalt Mixtures Containing Local Materials

For each performance test, existing research studies have likely already evaluated their sensitivity to asphalt mixture component properties or proportions (e.g., aggregates, asphalt binders, recycled materials, additives), volumetric parameters (e.g., air voids, VMA), and aging. However, contractors need to conduct their own analyses with their materials in order to make informed decisions on how to adjust their asphalt mixtures in the most cost-effective manner. One advantage promised by BMD is to give contractors the opportunity to be innovative with their asphalt mixture designs by evaluating new additives and combinations of materials to meet the performance test criteria in ways that are economically advantageous.

References

AASHTO R 9 (2018) *Standard Practice for Acceptance Sampling Plans for Highway Construction*, American Association of State Highway and Transportation Officials.

Ajede, A., Yin, F., and Tran, N. (2018) A Review of Acceptance Schedule of Payment for Asphalt Pavements in Alabama. Final Report of ALDOT Research Project 930-948.

Mohammad, L.N., M.A. Elseifi, S.B. Cooper III, C.S. Hughes, J.W. Button, and E.L. Dukatz, Jr. (2016) Comparing the Volumetric and Mechanical Properties of Laboratory and Field Specimens of Asphalt Concrete. NCHRP Report 818, ISBN 978-0-309-37508-5 | DOI 10.17226/23475.

Sholar, G., Page, G., Musselman, J., Upshaw, P., and Moseley, H. (2003) Development of the Florida Department of Transportation's Percent Within Limits Hot-Mix Asphalt Specification. FDOT Research Report FL/DOT/SMO/03-467.

Task 5 Checklist Establishing Baseline Data

- ☐ Analysis of pavement management system data and historical results with performance tests, if available, to assess the test applicability and help establish preliminary criteria.
- ☐ Plan the benchmarking study.
    - Determine how many mixtures will be collected and tested from existing mix categories.
    - Determine which labs will perform the tests for the benchmarking study.
    - Determine how to handle mix conditioning (lab aging) for cracking tests on surface mixtures.
- ☐ Execute benchmarking study plan.
- ☐ Analyze benchmarking data to estimate the distribution of results for current mixtures in each category and to determine the key mix factors that influence performance test results.
- ☐ Plan and execute the shadow projects using the new performance tests.
- ☐ Analysis of shadow project data and determination of the within-lot pooled standard deviations for the performance tests.
- ☐ Conduct testing and analysis by contractors to learn how to adjust and optimize their mixtures to meet preliminary criteria.

Chapter 6. Specifications and Program Development

Objective: To develop a performance specification and Quality Assurance Program for implementation.

Resources:

- AASHTO R 9, *Standard Practice for Acceptance Sampling Plans for Highway Construction.*

Outcome: (1) Refinement of the agency's sampling and testing plan including sublot size and assessment of agency and contractor risks. (2) Development of specifications and policies for pilot projects. (3) Execution and analysis of pilot projects and refinement of the specifications and QA program.

Task 6 involves the development of the specifications and the Quality Assurance Program that will be needed for implementation. The information gathered from the previous tasks will be used for this effort, such as identifying and validating the performance test selected (Subtasks 3.2 and 3.3), variability studies (Subtask 4.5), and development of baseline data (Task 5) to establish performance test criteria. Furthermore, information from the state DOT's existing QA program can be used to select the appropriate quality measures, Acceptance Quality Characteristics (AQCs), and preliminary specification limits for each test method. In this task, risk analyses can also be used to evaluate the QA program.

Based on the goals set by a state DOT (Subtask 2.4), there are several options of how acceptance and quality control testing can be handled for acceptance during mixture production. Examples are shown in Table 6.1.

Table 6.1. State DOT Examples for Asphalt Mixture Acceptance during Production.

State DOT	Acceptance Testing
Caltrans, LADOTD	Volumetric properties
NJDOT, TxDOT	Surrogate performance tests
IDOT, NJDOT, MaineDOT	Same performance tests as used for mixture design approval
NJDOT	Performance tests with pay adjustment factors

In some instances, the use of a surrogate performance test can be used in lieu of a more traditional performance test. The benefits for using surrogate performance tests include minimal investments by both the state DOT and industry, more tests can be completed within normal working hours, and reduced overall need for manpower and quick turnaround time. As an example, the Alabama DOT is considering using the High Temperature IDT (Hot IDT) test as a surrogate test for the Hamburg test, which would be used at design. However, surrogate tests for acceptance may require correlation/calibration with more fundamental performance tests.

6.1 Sampling & Testing Plans

Prior to developing a sampling and testing plan and associated QA program, a state DOT should review AASHTO R 9, *Acceptance Sampling Plans for Highway Construction*, which outlines the basic steps involved in this process. Prior to developing the plan, the following points should be considered as described in AASHTO R 9:

- Determine if outside expertise is needed, particularly if there is a lack of statistical knowledge or familiarity with the performance test.
- Review how other agencies have developed their specifications and QA programs, particularly with respect to the selection of quality measures and acceptance quality characteristics (AQCs).

In the development of the sampling and testing plan, consideration needs to be given to identifying the sampling location and methods, and defining appropriate lot and sublot sizes, including the impact lot and sublot sizes have on the reliability of acceptance decisions. Often the decision of sublot size, and therefore the sampling frequency, is based on logistical factors that affect the time needed to obtain results for each test. This includes the use of initial production lots to verify the key properties of the asphalt mixture design and the possible use of surrogate tests or screening tests in the QA program. While many agencies may choose to use their existing sampling locations as well as their current lot and sublot sizes, consideration should be given to the overall turnaround time for the selected performance test.

The lot size can have a significant economic impact on both the agency and the contractor, especially if there are delays in obtaining test data used to assess the lot. In general, the more samples taken from the lot, the more representative the data will be of the entire population. However, larger lot sizes can lead to greater economic risk for contractors as it represents more material placed prior to an acceptance decision, and typically contractors will offset this risk with higher bid prices. For example, if a contractor has a plant capable of producing 400 tons per hour and a performance sample is taken on day one and results are not available until day three, there could be close to 10,000 tons of material placed on the roadway before a decision is made on the acceptance of the material. This represents roughly a $750,000 risk for the contractor. So, while larger lot sizes may be more robust from a statistical perspective, smaller lot sizes can limit the financial risk for both the contractor and agency.

6.2 Pay Adjustment Factors (If Part of the Goals)

Adjustments to the pay for mixtures based on results of the selected AQCs is generally considered to be an effective means to encourage quality and ensure that the expected performance is achieved. The underlying assumptions for using pay adjustments are: (1) that the selected AQCs have a strong relationship to the pavement layer's resistance to a particular distress, and (2) the adjustments in pay for each AQC represents a fair compensation to the agency and the contractor for the positive or negative change in service life for the pavement. Evidence to support the first assumption for the new tests is growing as more field performance studies are published. However, evidence to support the second assumption has always been harder to establish given the interaction of numerous other variables on pavement performance such as underlying layer conditions and the adequacy of the pavement structural design for the actual traffic served on the roadway.

Burati et al. (2003) proposed the following equation to determine pay adjustment factors.

$$PAYADJ = \frac{C\left(R^D - R^E\right)}{\left(1 - R^O\right)} \qquad \text{Eq. 6.1}$$

where:

PAYADJ = appropriate payment adjustment for new pavement or overlay (same units as C),
C = unit cost of mixture,
D = design life of pavement layer,
E = expected (predicted) life of pavement layer,
O = Expected life of successive overlays,
R = (1 + INF)/(1 + INT),
INF = long–term annual inflation rate in decimal form, and
INT = long–term annual interest rate in decimal form.

The most challenging variable to determine in this equation is the expected (predicted) life of the pavement layer. Models for predicting the life of pavements or pavement layers have not been developed using current BMD index test results as direct inputs. The appropriate inflation and interest rates can also be points of disagreement among stakeholders. Sensitivity analyses to the inputs is recommended to facilitate discussions and help the decision-making process.

Currently, the most common approach to pay adjustment factors is based on AQC results for each lot. Some highway agencies use a pay factor equation and other use a pay factor table or schedule. These approaches should work well for AQCs based on BMD tests results as well.

Composite pay factors are used by many states to combine the pay adjustments from several AQCs. Generally, it is not recommended to combine pay factors from more than four AQCs as more than four will tend to diminish the impact of each AQC. Another approach is to apply only the lowest pay factor of the multiple AQCs. This is referred to as the “weakest link” approach suggesting that the AQC that is farthest outside of the specification limits will control the performance of the material.

The most important considerations in establishing fair pay factors are: (1) selecting AQCs that are related to performance, and (2) to set criteria for the AQCs that are realistically achievable.

6.3 Developing Pilot Specifications and Policies

A state DOT needs to develop specifications and revise other elements of its QA program that will be used for upcoming pilot projects (Subtask 6.4). This includes a compilation of the required acceptance tests, quality measures, AQCs, specification limits, sampling and testing frequencies, QC requirements, methods of validating contractor QC data (if contractor data is used for acceptance), and payment determination (if pay factors are part of the goals). In addition, a state DOT will need to address potential revisions to its independent assurance program and materials testing dispute resolution system (if contractor QC data is used for acceptance). These revised documents are to be used for pilot projects that will be constructed prior to full initial implementation.

In order to develop an acceptance plan for asphalt, the following items should be addressed:

- Quality Control (QC) requirements
- Acceptance requirements
- Independent Assurance requirements

- Agency validation (if contractor's tests are used in the acceptance decision)
- Materials Dispute Resolution system

Quality Control Requirements: Except in states that use contractor results in acceptance decisions, the majority of the existing QC requirements for the contractor will typically remain the same. However, some contractors may choose to obtain the appropriate testing equipment and run the performance tests for quality control or process control purposes, or they may possibly identify an appropriate surrogate test to run for process control. If QC tests are used in the acceptance decision, QC staff will be required to acquire the equipment and run the appropriate performance tests themselves, in addition to any other surrogate tests they may choose. Consequently, QC staff will need to be properly trained and their equipment assessed through the agency's Independent Assurance program.

Acceptance Requirements: At this point in the implementation process, one or more performance tests will have been selected, preliminary training will have been conducted on the test equipment and procedures, and benchmarking data will have been acquired. In addition, data analysis will provide much needed information on the variability associated with each test. All of this information will play a role in developing preliminary acceptance criteria that will be used in developing specifications for the pilot projects. Furthermore, information from the state DOT's existing QA program can be used when determining the most appropriate quality measures and acceptance quality characteristics for the specifications.

The goal of a performance specification is to link the design and acceptance characteristics of materials and construction to the expected performance of the pavement. This is typically accomplished by using a pay schedule that relates certain test results to the expected life of the pavement layer. The selected test results may include some traditional asphalt mixture properties as well as the results of performance tests or surrogate tests. Appropriate pay adjustments should be set to encourage the contractor to produce the materials as consistently as possible and be realistically achievable based on production variability data determined during pilot projects. The rationale for the pay adjustment schedule needs to be documented and well supported.

One determination that will need to be made relatively early in the process is who will be conducting the performance testing and where that testing will occur. Testing could be conducted by the agency as an acceptance test or by the contractor as a QC test or possibly as both. If the testing is a part of QC it must be determined if the data will be used in the acceptance decision. Testing location options include sampling the material at the asphalt plant or roadway and transporting it to a state DOT lab (or consultant lab) for acceptance testing (note that this could add additional time to the testing process) or requiring the contractor to obtain the testing equipment and have the material tested on-site for either acceptance or QC. One consideration is that until a final decision is made regarding the full adoption of performance requirements, it might be unreasonable to have contractors purchasing the testing equipment for pilot projects.

The next consideration is to determine which test results will be used to determine the acceptability of the asphalt mixtures. These test results are referred to as acceptance quality characteristics (AQCs). The selected AQCs may include just the performance tests identified in Subtask 3.2, or they may be a combination of the more traditional asphalt mixture properties testing (air voids, binder content, in-place density, etc.) as well as results of performance tests. Considering the general lack of familiarity contractors and project personnel will have with the new testing requirements, a conservative approach would be to retain one or two of the existing AQCs and use them in conjunction with the performance test. However, it is important that the AQCs selected not conflict with each other. For example, if a cracking test is selected, it may prove challenging for the contractor to meet the cracking test criteria and also meet the specified criteria for production air voids or binder content. It is also worth noting that there is a practical limit to the number of AQCs that should be used for acceptance and pay factor determination for a given material. In general, more than four AQCs can dilute the impact of any individual characteristic.

Following the selection of the AQCs, the next step will be to determine which Quality Measure should be used to relate the testing results to payment. The most common quality measure in use today is Percent Within Limits (PWL), which is used by 31 state highway agencies for asphalt mixture acceptance (Dvorak 2021). Other quality measures currently in use today (and the percentage of states that use it) include Average Absolute Deviation (10%), Single Test (8%), and Single Test and Running Average (8%). While there are advantages and disadvantages to each approach, consideration should be given to adopting a Quality Measure that addresses both accuracy to the target value and variability, such as PWL. While AAD also addresses accuracy and variability to a certain extent, AAD specifications require both minimum and maximum limits, whereas most of the performance tests in use today only have minimum or maximum criteria, not both. A limitation of PWL specifications is that it requires test results obtained from at least three sublots, and, as discussed later in this section, using a small number of sublots increases risks to the agency and contractor. Therefore, when a limited quantity of a mixture is used on a project (e.g. less than 3000 tons) an alternate method should be used to determine acceptance of the mixture that is based on one to three test results.

If PWL specifications are used, the first step will be to determine the Acceptable Quality Level (AQL), and the Rejectable Quality Level (RQL). Material that meets the AQL typically has a payment of 100% (or more), while material that fails to meet the RQL is typically rejected or, in some cases, left in place at substantially reduced pay.

The AQL is the minimum level of quality at which the material can be considered fully acceptable (for that acceptance characteristic), which would then receive a payment of at least 100%. Fundamentally, this is an engineering decision based on knowledge of past performance, and generally reflects material that will provide the expected service life. While many agencies use an AQL of 90 PWL, other PWL levels are also commonly used. AQLs are typically developed in conjunction with the acceptance characteristic specification limits (described below).

The RQL is the minimum required level of quality for the material. Levels below the RQL are typically either rejected or left in place at substantially reduced pay. Establishing the RQL can be based on an engineering decision related to how much material is allowed to fall outside of the specification limits, performance of past projects, and costs to repair the deficient pavement. An RQL of 60 PWL is common.

The next step is to determine the appropriate specification limits and targets for the AQCs, and for this there are several options available. Specification limits can be based on data correlated with actual field performance (Subtask 3.4), they can be based on values from another state or a national study, they can be based on typical values determined from the benchmarking and shadow projects, or a combination of these options.

The recommended approach for establishing specification limits and targets is to use data that has been correlated with actual field performance. However, it is unlikely long-term performance data will be available for the pilot projects. Another option would be to use target values and limits developed from a field experiment in another state or national study. While this may be a relatively simple solution, caution should be exercised to make certain that the values used are realistic for the materials and construction practices used locally. The most practical method of determining preliminary specification limits may be to use data collected from the benchmarking studies and shadow projects. As was discussed in Chapter 5, target values and variability can be established from this data. From the variability data coupled with the AQL and the specification limit, targets can then be determined. An example illustrating appropriate targets for a given set of criteria and variability data in a PWL specification is provided in Appendix C.

Estimating Risks: Statistically-based acceptance plans for construction materials are based on testing representative samples of the material to estimate the quality of a larger quantity (i.e., a lot) of the material. This approach involves risk – that is, there is a chance that the random samples are not representative of the lot, or testing errors cause an incorrect estimate of the lot's quality. Therefore, there are inherent risks to the agency (buyer) and the contractor (seller) as part of statistical acceptance plans:

- Seller's risk, known as the risk of a Type I error (α), occurs when acceptable quality is rejected as unsatisfactory and results in unnecessary removal and reconstruction of the lot.
- Buyer's risk, known as the risk of a Type II error (β), occurs when unacceptable quality is judged to be satisfactory and can result in additional pavement maintenance costs or premature failure.

It should be noted that a good acceptance sampling plan does not eliminate risks, rather it balances risks versus costs in an objective way. Operational Characteristic (OC) curves are a statistical tool to assess the risks of a given specification and sampling plan. OC curves allow stakeholders to easily see how changing criteria that define the required performance and the required amount of testing impact risks. Therefore, they are helpful to balance costs of testing and the risks of making Type I and Type II errors.

OC curves are defined by five parameters: (1) the Acceptable Quality Limit (AQL), (2) the Rejectable Quality Limit (RQL), (3) the seller's risk, (4) the buyer's risk, and (5) the number of samples. The AQL is the percent defective at which the lot still receives full payment. The RQL is the percent defective at which the lot is rejected. The seller's (contractor's) risk is the probability that the tests results will indicate that the lot is unacceptable when the lot is actually within in the specification limits. The buyer's (agency's) risk is the chance of accepting a lot based on test results when the lot of material is not actually within specification limits (i.e., saying the lot is good when it is actually bad). The number of samples is number of test results used to make the decision to accept or reject a lot.

One way to use OC curves is to input four of the five parameters into an OC curve analysis program that will output the fifth parameter. An example is provided in Appendix D. The main point to understand is that risk level is inversely related to the number of tests (i.e. sublots) used in the acceptance decision for each lot; more test results reduces Type I and II risks and vice versa.

Independent Assurance Requirements: Independent Assurance (IA) is an unbiased and independent evaluation of all the sampling and testing procedures used in the acceptance program. It evaluates the qualified sampling and testing personnel as well as the testing equipment. The IA program must cover sampling procedures, testing procedures, and testing equipment, and the program itself can be either Systems-Based or Project-Based. Under the new performance specifications, a state DOT's IA program would likely function the same as it does currently, and the new performance testing requirements would simply be added. For example, if an agency used a systems-based program that relied on proficiency samples as a method of evaluation, then samples would need to be distributed for the new performance tests. Conversely, if an agency used a project-based system that relied on observations and split samples, performance samples would need to be incorporated into this method, and appropriate observation checklists would need to be developed and inspection staff properly trained. One important item of note, until the inter-laboratory studies described in Section 4.5 are completed and d2s limits are established, allowable variability within- and between-labs will not be known, which could have an impact on running split samples.

Since pilot projects will include actual requirements that are contractually binding (including those for Independent Assurance), SHAs must develop a preliminary method to assess sampling and testing personnel involved in the performance testing, as well as the laboratory equipment. This would likely include at a minimum a checklist that could be used in observing the technician performing the test and also for ensuring that the equipment is properly calibrated and maintained. As more laboratories begin to routinely run the performance tests, and an interlaboratory study is conducted, a proficiency sample program and split-sample analysis will then be an option. The pilot projects will provide a valuable opportunity to assess which method of IA will be the most appropriate for full implementation.

Agency validation of contractor test results: The FHWA policy on the use of contractor's quality control test results for acceptance requires validation of all data not generated by the state DOT or its designated agent if used in the acceptance decision. If the performance tests selected are to be performed by contractor personnel and used in the acceptance decision, the data must be validated. While there are numerous methods currently used to validate QC data, many do not fully meet federal requirements, which include:

- The use of random samples for acceptance tests
- Validation samples are obtained independently from the contractor samples (not splits)
- All agency and contractor personnel involved in acceptance testing are qualified in accordance with the state's approved QA program.

The recommended method of validating contractor data with independent samples includes statistical comparisons of the contractor and agency data from the same quantity of material or area of construction using the F-test and the t-tests to compare variances and means, respectively.

- The F-test provides a method for comparing the variances (standard deviations squared, σ^2) of two sets of data to determine if they come from the same population. The analysis is conducted by assessing the size of the ratio of the variances.
- The t-test provides a method for comparing the means of two independent data sets and is used to assess the degree of difference in the means.

Materials Dispute Resolution system. If the results from the quality control sampling and testing are used in the acceptance program, the state DOT must have a dispute resolution system. The dispute resolution system addresses discrepancies that may occur between the verification sampling and testing results and the quality control sampling and testing results. This frequently involves testing split samples obtained from the Quality Control and Verification samples with an analysis involving either the d2s limits (which compares the contractor and department results from a single split sample) or possibly the use of a paired t-test (which compares contractor and department results from an equal number of split samples).

6.4 Conducting Pilot Projects

Pilot projects are necessary to evaluate the new QA program requirements under actual contractual conditions. The pilot projects will go through the typical bidding-contracting process with the new QA requirements and specifications applied, including performance testing required as part of asphalt mixture design and acceptance. As part of the pilot project process, state DOTs need to develop and conduct just-in-time training to address any necessary testing or specification changes associated with the QA program. The number of pilot projects necessary should be determined by the state DOT and can start out with just a few in the first year, then increase to involve additional districts and contractors in subsequent years as more experience is gained. Pilot projects would include meetings with technical support staff and mandatory pre-bid conferences to discuss the new requirements, along with discussions on how to address problems that may occur during their construction. The just-in-time training would likely be conducted by state DOT staff or researchers who were involved with performance testing that was conducted previously in order to utilize lessons learned and hands-on experience as the additional technicians are trained. The agency needs to be flexible with the pilot projects to allow the capability to make changes on future or on-going projects based on lessons learned from other on-going or completed projects.

At the completion of each pilot project, the technical support staff should meet with the contractor, project, and district personnel to review the project and discuss problems and concerns. This will help to determine the various issues that still need to be addressed and may also identify successes and challenges that can be shared with other on-going projects. Following the close-out meeting, a summary report should be prepared documenting the project.

One important consideration is the continued communication with the Stakeholders/Task Force (Section 2.2) on issues that are encountered on the pilot projects. For example, since the pilot projects will involve new testing and specification requirements, it is likely that bid prices will be significantly higher than what would be expected on a conventional project, or it is possible that the failure rates may be disproportionately too high or low. It is important that this information be communicated back to the Task Force along with an appropriate explanation that these are pilot projects and with time, bid prices and failure rates will adjust.

6.5 Final Analysis and Specification Revisions

A state DOT needs to conduct a comprehensive review of the specifications as well as any other necessary changes to their QA program based upon the lessons learned from the pilot projects. Data collected as part of the pilot projects should be closely analyzed and the specification requirements modified accordingly. In addition, any systemic problems that were encountered on the projects will also need to be addressed. This is also a good time to review the performance of field validation test sections, shadow projects, and other sources of information outside of the state and consider adjustments to the preliminary specification criteria. Specification revisions and any necessary changes need to be completed prior to initial implementation.

References

AASHTO R 9 (2018) *Standard Practice for Acceptance Sampling Plans for Highway Construction*, American Association of State Highway and Transportation Officials.

AASHTO R 42 (2016) *Standard Practice for Developing a Quality Assurance Plan for Hot Mix Asphalt (HMA)*, American Association of State Highway and Transportation Officials.

Burati, J.L., R. Weed, C. Hughes, and H. Hill (2003) Optimal Procedures for Quality Assurance Specifications, FHWA-RD—2-095, Federal Highway Administration.

Federal Highway Administration (2004) Use of Contractor Test Results in the Acceptance Decision, Recommended Quality Measures, and the Identification of Contractor/Department Risks, https://www.fhwa.dot.gov/construction/t61203.cfm

Hand, A.J.T, M. Nimeri, E. Hajj, P. Sebaaly, R. West, M. Heitzman, F. Yin. C. Hughes, and S. Tayabji (2020) Procedures and Guidelines for Validating Contractor Test Data. NCHRP Report 946. The National Academies Press. https://doi.org/10.17226/25823.

Transportation Research Board, Glossary of Transportation Construction Quality Assurance Terms (2018) Seventh Edition, Transportation Research Circular E-C235. ISSN 0097-8515.

Task 6 Checklist

- ☐ Review AASHTO R 9 and AASHTO R 42
- ☐ Determine if outside expertise is needed for developing the sampling and testing plan
- ☐ Select the sampling location for asphalt mixture samples used in acceptance testing
- ☐ Select the sublot sizes (mix quantity)
- ☐ Set the QC requirements
- ☐ Determine if Contractor data will be used in acceptance decisions
- ☐ Select the Acceptance Quality Characteristics
- ☐ Determine the Quality Measure used to assess lots (e.g. PWL)
    - Set Acceptable Quality Level (AQL)
    - Set Rejectable Quality Level (RQL)
    - Set specification limits for the AQCs (i.e. criteria for cracking and rutting tests)
- ☐ Use OC curves to evaluate risks and the minimum number of samples (sublots)
- ☐ Review and modify the Independent Assurance plan as needed
- ☐ Review and modify the validation procedure and policies if Contractor data is used in acceptance decisions
- ☐ Review and modify as needed the dispute resolution system
- ☐ Write and review the specifications and QA program for the Pilot Projects
- ☐ Plan the Pilot Projects
    - Determine the number, general locations, and types of projects
    - Conduct Training for agency and contractor personnel involved in Pilot Projects
- ☐ Execute the Pilot Projects
    - Provide additional resources as needed to conduct the new test
    - Analyze results
    - Review results and lessons learned
- ☐ Adjust the specifications and QA program for future projects

Chapter 7. Training, Qualifications, and Accreditations

Objectives: To plan the development or modification of the asphalt technician training and qualification program and the laboratory accreditation program.

Resources:

- FHWA Guidance on Quality Assurance: https://www.fhwa.dot.gov/pavement/0637bsup.cfm
- AASHTO Accreditation Program: http://www.aashtoresource.org/aap/overview
- AASHTO re:source Proficiency Sample Program: http://www.aashtoresource.org/psp/overview

Outcomes: (1) Update the training and qualification program to include instructions on new test methods and associated sample preparation procedures as well as other changes made to the state's QA program. (2) Update the requirements for sample preparation and performance testing equipment to the state's laboratory accreditation program. (3) (Optional) Establish or modify the state's proficiency sample program to include the new performance tests.

Following completion of the pilot projects and prior to the full statewide implementation of the performance specifications, a state DOT will need to formalize the changes to its existing technician qualification and laboratory accreditation programs, in order to assure that all testing meets the requirements of 23 CFR 637 B.

7.1 Developing and/or Updating Training and Certification Programs

State DOTs will need to update their existing technician certification programs to include the new sampling and testing requirements associated with the adoption of the performance specifications. However, this is somewhat of a long-term solution since it may take several years to get all currently certified technicians "recertified". Consequently, the agency will need to address how training will be provided for personnel already certified for the previously used acceptance tests.

One short-term solution to this problem is to use district- or region-wide workshops to provide training on the initial projects until the new tests and requirements can be incorporated into the formal programs. Course manuals that address the new sampling and testing requirements can be developed and provided to workshop participants and should include detailed descriptions and photos of the test methods, including equipment, sampling, conditioning, specimen preparation, test procedures, etc. Having instructional videos highlighting the details for sample preparation, specimen fabrication, testing, and data analysis will be extremely helpful for technicians and other personnel involved in the implementation process of the performance tests. Offering one-on-one support is also an effective tool.

The course materials developed for the workshops can then be used to modify the existing certification programs, providing a long-term solution.

7.2 Establishing or Updating Laboratory Accreditation Program Requirements

With respect to laboratory accreditation, since many of the asphalt tests used in performance testing are relatively new, current accreditation standards do not exist since the national accreditation programs typically require that a test method has a fully approved and accepted standard. While most of the cracking and rutting tests have AASHTO or ASTM standards, many are still in the provisional stage. A state DOT will ultimately need to add the performance tests to its own (or regional) laboratory accreditation program. Until the laboratories are fully accredited, performance test equipment checks will need to be added to the state DOT's routine checklist for all laboratories conducting the tests, especially if they are used in the acceptance decision. Critical equipment/calibration checks are needed to assist a state DOT in updating its inspections by adding these test procedures.

7.3 Establishing a Proficiency Sample Program

A state DOT can also establish and implement a statewide proficiency sample program for all technicians involved in performance testing of asphalt mixtures for design and acceptance. This involves technicians fabricating and testing standardized specimens and reporting results for analysis. A proficiency sample program ensures that technicians are properly performing the tests in accordance with applicable standard methods, provides useful updates to repeatability and reproducibility data, and also identifies which laboratories may have results that are significantly different from the overall population. Such laboratories would need to carefully review their procedures for sample preparation and conducting the test.

A state DOT can also maintain a type of material producer list (MPL) for all laboratories approved to perform the performance tests. For example, the approval process involves an initial split sample testing with the state DOT laboratory and participation in an annual state-wide performance testing proficiency program (e.g., TxDOT).

Chapter 8. Initial Implementation

Objectives: To present guidance on transitioning the new methods, specifications, and QA program to full implementation.

Outcomes: (1) Communication of changes to all stakeholders. (2) Establish a process for feedback and to periodically review and update the test methods, specifications, and QA program as new information becomes available.

Prior to full implementation of the performance requirements, it is essential that the state DOT adequately communicate the changes and new requirements to both industry and agency personnel. This technology transfer can be accomplished through webinars, face-to-face meetings, and workshops. It can also be supported by having “implementation teams” to help contractors and state DOT personnel address problems, interpret specification requirements, etc.

Project selection guidelines, as discussed in Subtask 2.4, may need to be revised prior to full implementation of the performance specifications. The initial scope of projects for which performance testing is used may be revisited with consideration of the project investment and risk levels, the type of pavement rehabilitation, the project length and asphalt tonnage, the pavement layer, and the resources required for implementation.

It is important to integrate a feedback loop into the process to ensure and encourage communication and regular feedback from the various stakeholders, to share new information as it becomes available, and to help identify areas for adjustment and improvement. Feedback loops help a state DOT have a more coordinated, collaborative, and committed effort towards full implementation of BMD and performance specifications. This involves continuous monitoring of test sections and early projects that were built as part of this overall BMD implementation effort.

Appendix A. Procedure for Evaluating Potential Outliers from a Lot

This procedure was developed by the Maine Department of Transportation. It is adapted from ASTM E 178 *Dealing with Outlying Observations* and was recommended in NCHRP Report 946.

Scope: This procedure deals with identifying outlying observations in sets of at least three results.

Definition: An outlying observation, or "outlier," is one that appears to deviate markedly from other test results from the same population or lot. When considering outliers, two conditions may exist: 1) the value may be an extreme value of the population or excessive variability of the population, and in either case, the value should not be discarded, or 2) it may be the result of gross deviation from prescribed sampling and or test procedures, errors in calculations, or errors in recording of numerical values, in which case it should be discarded. The procedure below provides steps to determine which of the two decisions to make, i.e., the value is not an outlier and should be retained, or the value is an outlier and should be discarded.

Procedure:

Step 1. Determine if a physical reason is known for the outlier. Possible reasons may include: the sample was mishandled prior to testing, the test equipment malfunctioned, or a computation error was made. If a computation error is found, it may be corrected, and the corrected value used as the test result.

Step 2. If no reason is found for the outlier, the following calculation procedure should be used.

Calculation Procedure - This procedure is based on a "two tail *t*-test" with level of significance (α) of 5%. The two-tail test means that the outlier may be either on the high or low side of the average. The level of significance means that if a value is identified as an outlier, there is only a 5% chance that it is not.

1. Calculate the sample average ($\overline{x}$) and standard deviation (s) of the results in the sample set (e.g., lot).
2. Find the critical t value 't_{crit}' from Table A.1 using the total number of samples (n) in the sample set.
3. Determine D, the total allowable deviation on either side of $\overline{x}$, by multiplying 't_{crit}' by s.
4. Establish values for MAX and MIN by adding and subtracting D to and from $\overline{x}$.
5. Any result greater than MAX or less than MIN is determined to be an outlier.

Example 1. The following eight (8) density test results were obtained. Is one an outlier?

Sample	Relative Density, %
1	89.5
2	94.0
3	93.3
4	93.3
5	92.8
6	92.6
7	93.5
8	94.3

Calculations

n = 8

$\overline{x}$ = 92.9

s = 1.49

t_{crit} = 2.126 (from Table A.1)

D = $t_{crit} \times s$ = 2.126 × 1.49 = 3.17

MAX = $\overline{x}$ + D = 92.9 + 3.17 = 96.07%

MIN = $\overline{x}$ – D = 92.9 – 3.17 = 89.73%

Since Sample 1 is 89.5%, which is less than the MIN of 89.73%, this sample result is identified as an outlier and should be investigated.

Example 2. The following three air void results were obtained. Is one an outlier?

Sample	Air Voids, %
1	5.2
2	5.2
3	6.6

Calculations

n = 3

$\overline{x}$ = 5.7

$s = 0.81$

$t_{crit} = 1.155$ (from Table A.1)

$D = t_{crit} \times s = 1.155 \times 0.81 = 0.94$

MAX = $\overline{x} + D = 5.7 + 0.94 = 6.64\%$

MIN = $\overline{x} - D = 5.7 – 0.94 = 4.76\%$

Since Sample 3 is 6.6%, which is less than MAX of 6.64%, this sample result is not an outlier and should be used in further calculations.

TABLE A.1 t_{crit} values for a 5% Significance Level

n	t_{crit}
3	1.155
4	1.481
5	1.715
6	1.887
7	2.020
8	2.126
9	2.215
10	2.290
11	2.355
12	2.412
13	2.462
14	2.507
15	2.549
16	2.585
17	2.620
18	2.651
19	2.681
20	2.709
21	2.733
22	2.758
23	2.781
24	2.802
25	2.822
26	2.841
27	2.859
28	2.876
29	2.893
30	2.908

Appendix B. Example Normality Check of Flexibility Index Data from a Shadow Project

Figure B.1 shows a probability plot for 14 results of Flexibility Index (FI) from a single project with the Anderson-Darling test for normality in Minitab statistical software. In this case, the small Anderson-Darling (AD) statistic and p-value > 0.05 indicates that the sample data can be assumed to come from a normal distribution. The relative alignment of the FI data along the probability line is also a visual indicator of the normality check. Figure B.2 shows a histogram of the same data.

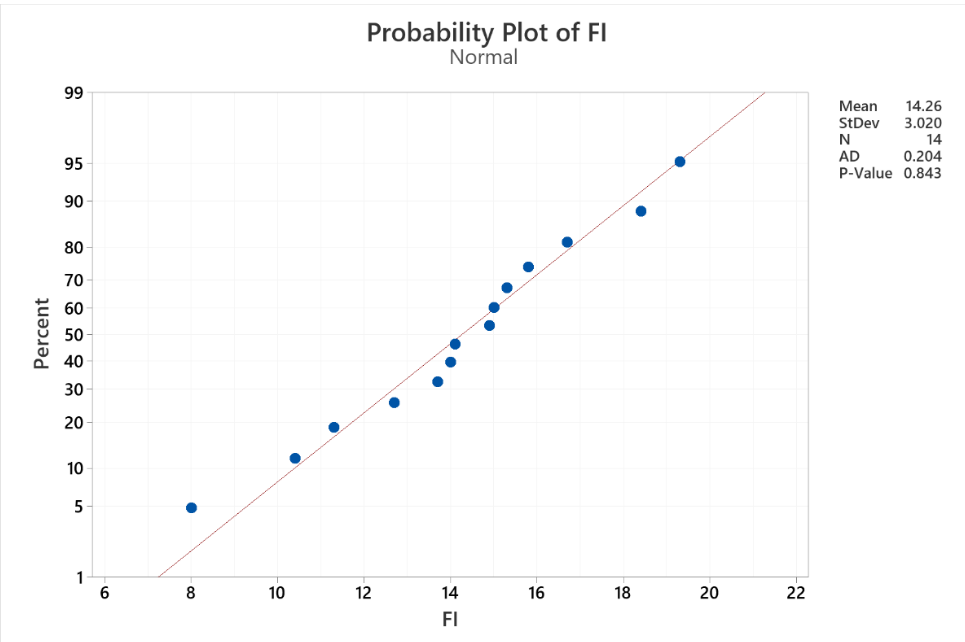


Figure B.1. Probability plot of FI from a project with an Anderson-Darling Normality Test.

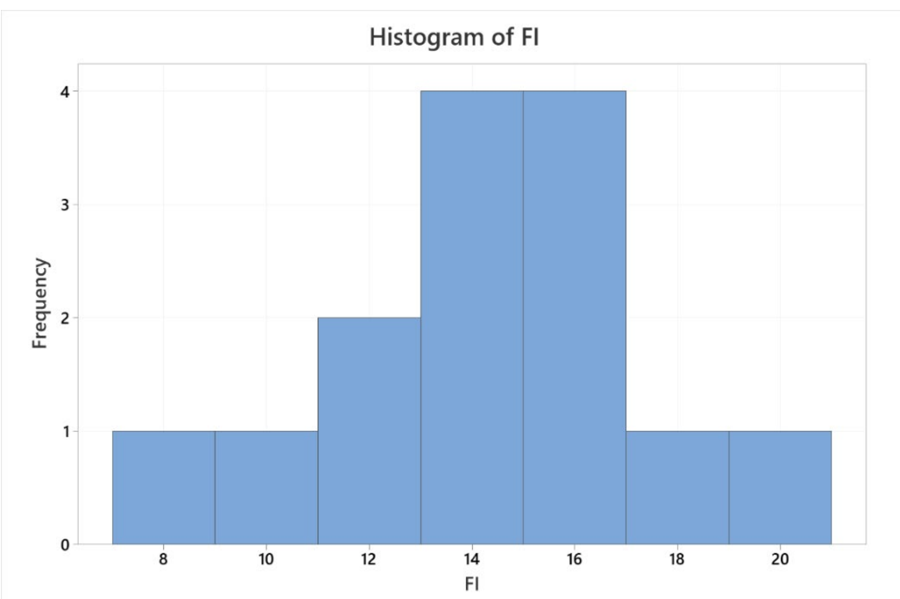


Figure B.2. Histogram of Flexibility Index results from a single project for visually assessing normality.

Appendix C. Example Using Performance Tests in a Percent Within Limits Specification

For this example, an agency uses a PWL specification for acceptance of asphalt mixtures and defines the acceptable quality level (AQL) as 95 PWL, and the rejectable quality level (RQL) as 60 PWL. The agency typically uses four test results to represent one lot of material. Based on the standard "Quality Index Value Table" shown in Table C.1, for n = 4, the quality indices corresponding to the AQL and RQL are 1.35 and 0.30, respectively.

Table C.1. Partial Quality Index Table for Estimating PWL.

PWL	*n* = 3	*n* = 4	*n* = 5	*n* = 6
100	1.16	1.50	1.79	2.03
99	–	1.47	1.67	1.80
95	–	1.35	1.44	1.49
90	1.10	1.20	1.23	1.24
85	1.03	1.05	1.05	1.04
80	0.93	0.90	0.88	0.87
75	0.82	0.75	0.72	0.71
70	0.68	0.60	0.57	0.56
65	0.52	0.45	0.43	0.41
60	0.36	0.30	0.28	0.27
55	0.18	0.15	0.14	0.14
50	0.00	0.00	0.00	0.00

Source: Specification Conformity Analysis, FHWA Technical Advisory T5080.12, June 23, 1989.

The agency is considering the IDEAL-CT test for acceptance with a minimum specification limit for CT_{Index} of 40. This is a single-sided specification limit such as might be used for any index test. Therefore, the lower quality index, Q_L, is calculated using Equation C.1.

$$Q_L = \frac{(\bar{X} - LSL)}{s} \qquad \text{Eq. C.1}$$

Where $\bar{x}$ is the lot average; LSL is the lower specification limit; and s is the lot standard deviation.

From the shadow projects, the estimated typical within-lot standard deviation for CT_{Index} was 9.2. With this information, the lowest target that a contractor should use for CT_{Index} to achieve full pay can be determined by rearranging Equation C.1 to solve for $\bar{x}$,

$$\bar{x} = (Q_L \times s) + LSL \qquad \text{Eq. C.2}$$

Full pay (AQL = 95 PWL, Q_L = 1.35): $\bar{x} = (1.35 \times 9.2) + 40 = 52.4$

Thus, contractors in this state should target a CT_{Index} greater than or equal to 52.4 and/or a standard deviation less than or equal to 9.2 to achieve full pay. Similarly, the target CT_{Index} that risks the lot being completely rejected is determined as:

Reject (RQL < 60 PWL, Q_L = 0.30): $\bar{x} = (0.30 \times 9.2) + 40 = 42.8$

Of course, if a contractor can produce mix with a smaller standard deviation than the "typical value" used in the above calculations, then he/she could have lower target values; conversely, if the contractor's standard deviation is greater that the above calculations, then his/her targets would have to be higher.

Thus, the range of within-lot production variabilities from the shadow projects will provide insight as to what is good versus poor variability in the state and contractors can estimate their risks accordingly.

Appendix D. Example Analysis of a Sampling Plan Using Operational Characteristic Curves

This example illustrates the use of an Operational Characteristic (OC) curve analysis for determining the minimum number of samples (i.e. sublots) used to make an acceptance decision. This example uses the OC curve tool in the Minitab statistics program. For highway construction materials, percent within limits (PWL) is the most common quality measure to estimate how well the lot of material meets the specification based on the mean and standard deviation of test results conducted according to the agency's acceptance sampling plan. In this example, the term "Percent Defective" is used since that how the OC tool in Minitab was developed. Percent Defective (PD) is simply the complement to PWL as shown in the following equation.

$$PD = 1 - PWL$$

Figure D.1 shows an OC curve based on the following four inputs: (1) an AQL of 10% percent defective, (2) a RQL of 30% defective, (3) seller's risk of 0.05 (i.e. a 5% chance of making the wrong decision), and (4) buyer's risk of 0.05. The resulting OC curve plot shows the percent defective (x-axis) versus the probability that the lot will be accepted (y-axis). For this set of conditions, the minimum sample size is 19. This means that for an acceptable quality level of 90% PWL, a rejectable quality level of 70% PWL, and balanced agency and contractor risks of 5%, then the sampling plan would require a minimum of 19 sublots. This is number of sublots considered impractical for typical highway paving projects.

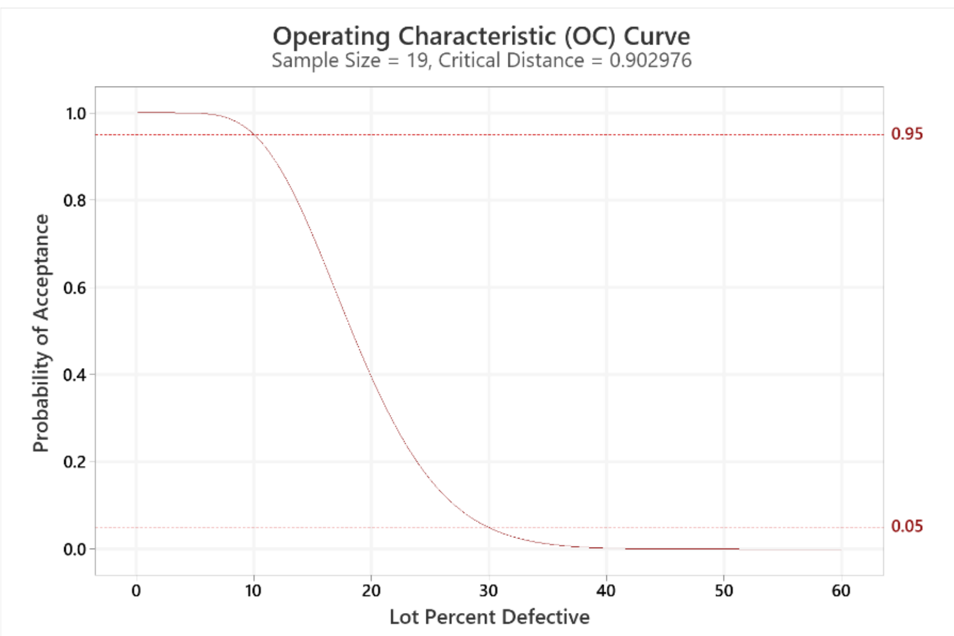


Figure D.1. Example Operational Characteristic Curve.

To obtain a smaller sample size, there are several options:

1. Increase the contractor's risk, α
2. Increase the agency's risk, β
3. Lower the AQL for PD (or raise the AQL for PWL)

4. Raise the RQL for PD (or raise the RQL for PWL)
5. Some combination of the above.

To illustrate options 3 and 4 simultaneously, lowering the PD AQL to 0.05 and raising the PD RQL to 40 results in the OC curve shown in Figure D.2, which changes the minimum number of samples to 6 which is a more reasonable number of tests (sublots).

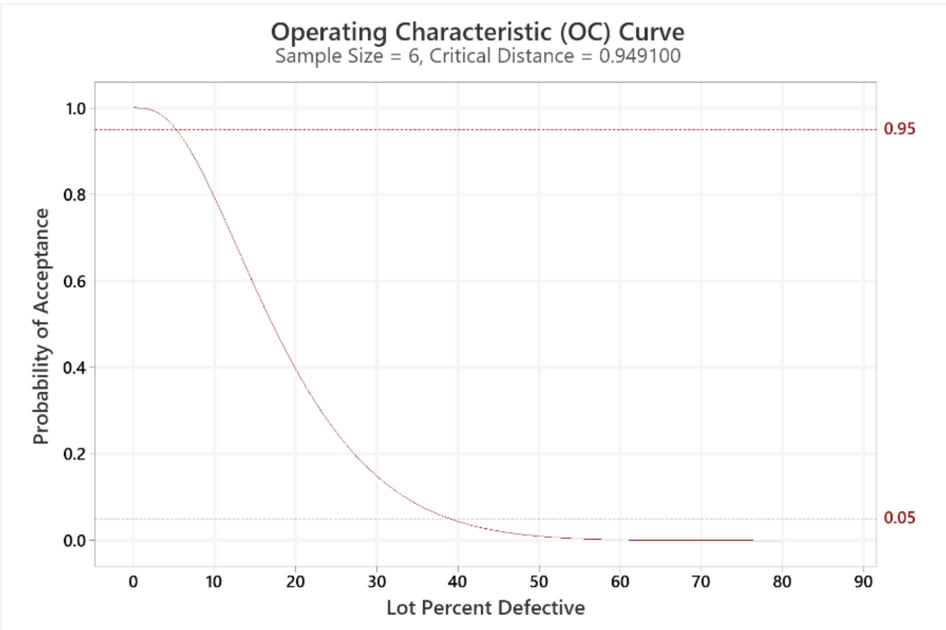


Figure D.2. Example Operational Characteristic Curve.