Technical Report Documentation Page

1. Report No.
FHWA-RD-02-095

2. Government Accession No.

3. Recipient's Catalog No.

4. Title and Subtitle
Optimal Procedures for Quality Assurance Specifications

5. Report Date

6. Performing Organization Code

7. Author(s)
J. L. Burati, R. M. Weed, C. S. Hughes, H. S. Hill

8. Performing Organization Report No.

9. Performing Organization Name and Address
Clemson University
Civil Engineering Department
Clemson, SC 29634-0911

10. Work Unit No. (TRAIS)

11. Contract or Grant No.
DTFH61-98-C-00069

12. Sponsoring Agency Name and Address
Office of Research, Development, and Technology
Federal Highway Administration (FHWA)
6300 Georgetown Pike
McLean, VA 22101-2296

13. Type of Report and Period Covered
Final Report

14. Sponsoring Agency Code

15. Supplementary Notes

This study was conducted under State Planning and Research Pooled-Fund Study number 2(199) and was administered by FHWA. FHWA Contracting Officer's Technical Representative (COTR): Peter A. Kopac, HRDI-12.

16. Abstract

This manual is a comprehensive guide that a highway agency can use when developing new, or modifying existing, acceptance plans and quality assurance specifications. It provides necessary instruction and illustrative examples to lead the agency through the entire process of acceptance plan development, including:

- Setting up the initial data collection/experimentation to determine typical parameters of current construction.
- Establishing the desired level of quality to be specified.
- Designing the actual acceptance plan itself, including selecting quality characteristics, statistical quality measure, buyer's and seller's risks, lot size, number of samples (sample size), specification and/or acceptance limits, and payment–adjustment provisions.
- Monitoring how the acceptance plan is performing.
- Making necessary adjustments.

17 Key Words
Quality Assurance, Quality Control, Specifications, Statistical Specifications, QA, QC, Payment Adjustments

18. Distribution Statement
No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19 Security Classif. (of this report)
Unclassified

20. Security Classif. (of this page)
Unclassified

21. No. of Pages
347

22. Price

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: volumes greater than 1000 L shall be shown in m^3				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
^{o}F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	^{o}C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
^{o}C	Celsius	1.8C+32	Fahrenheit	^{o}F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

OPTIMAL PROCEDURES FOR QUALITY ASSURANCE SPECIFICATIONS

PREFACE

This manual is intended as a comprehensive guide that a highway agency can use when developing new, or modifying existing, acceptance plans and quality assurance (QA) specifications. It is intended to provide necessary instruction and illustrative examples to lead the agency through the entire process of acceptance plan development, including:

- Setting up the initial data collection/experimentation to determine typical parameters of current construction.
- Establishing the desired level of quality to be specified.
- Designing the actual acceptance plan itself, including selecting quality characteristics, statistical quality measure, buyer's and seller's risks, lot size, number of samples (sample size), specification and/or acceptance limits, and payment–adjustment provisions.
- Monitoring how the acceptance plan is performing.
- Making necessary adjustments.

The acceptance plans described in this manual are based on QA principles intended to provide the *"…confidence that a product or facility will perform satisfactorily in service."* (2) To achieve this end, the acceptance plans must be realistic, must be fair to both the contractor and agency, and must be statistically accurate. The purpose of this manual is to explain how this can be done, through discussion and with examples where appropriate.

The information in this manual is based on a pooled–fund study that was administered by the Federal Highway Administration (FHWA). The material presented herein is based on review of the literature, the input of a panel comprised of a representative from each of the States in the pooled fund, numerous statistical analyses conducted for the project, and the past experiences of the authors. Additional information and more detailed discussion, as well as a thorough summary of all supporting analyses, can be found in the technical report for the project. (17)

TABLE OF CONTENTS

LIST OF FIGURES

Figure **Page**

LIST OF TABLES

LIST OF EQUATIONS

Equation **Page**

CHAPTER 1. INTRODUCTION

Historical Perspective

Quality Assurance (QA) acceptance plans are being used or developed by about 90 percent of State Highway Agencies (SHAs) and most Federal transportation agencies. This has been an ongoing, evolutionary process. This evolution has taken place over several decades and has led to much–improved acceptance plans over those used in the past. The genesis of the move toward QA began in 1956 with the American Association of State Highway Officials (AASHO) Road Test, and the analyses that emanated from that historic study. (1) The unsuspected discovery of the large magnitude of the variability in materials and construction led to the conclusion that specifications must be improved. This, in turn, led to the first step in the evolution of QA specifications and acceptance plans. Prior to the AASHO Road Test, with few exceptions, specifications were materials and methods specifications, sometimes called "prescription" or "recipe" specifications.

The first forms of QA specifications were called statistically oriented specifications or statistically based specifications. These evolved into more modern QA specifications that stress the need to separate quality control (QC) from acceptance. The evolution is now emphasizing the need for performance–related specifications (PRS) that not only describe the desired levels of selected quality characteristics, but also employ quantified relationships containing these characteristics to predict subsequent pavement performance. (2)

The description of the evolution of QA and PRS specifications is well documented in *National Cooperative Highway Research Program* (*NCHRP) Synthesis 38*, "Statistically Oriented End–Result Specifications," (3) *NCHRP Synthesis 65*, "Quality Assurance," (4) and *NCHRP Synthesis 212*, "Performance–Related Specifications for Highway Construction and Rehabilitation." (5) These syntheses provide an excellent history for the reader interested in how and when QA specifications evolved.

Important Definitions

There are two particularly important sources for identifying definitions and terminology associated with QA specifications. The first of these is the Transportation Research Board (TRB) *Transportation Research Circular Number E–C037*, "Glossary of Highway Quality Assurance Terms," (2) which provides a comprehensive glossary of definitions and terminology. This document is referred to in this manual as the "TRB glossary." The second important source for definitions and terminology is the FHWA Federal–Aid Policy Guide, *23 CFR 637B*, "Quality Assurance Procedures for Construction." (6) This document is referred to in this manual as "FHWA *23 CFR 637B*."

It is very important to have a clear understanding of the terms used in this manual. To facilitate this understanding, the definitions included in the TRB glossary (2) are used whenever possible. When other sources of definitions are used that differ from those in the TRB glossary, an explanation of the difference is provided.

Definitions Relating to Quality

The TRB glossary (2) includes the following definition for quality:

- ***Quality****—(1) The degree of excellence of a product or service; (2) the degree to which a product or service satisfies the needs of a specific customer; or (3) the degree to which a product or service conforms with a given requirement.*

This definition really indicates that there are three different definitions of quality. These are often referred to as the "level of goodness," "customer satisfaction," and "conformance to requirements" definitions. From a contractor's perspective, the conformance to requirements definition is the most appealing since the contractor's primary role is to satisfy the specification requirements. However, if the specification is not written in the proper manner, meeting the specification requirements may not meet the customer's expectations and therefore may not satisfy the customer. It is therefore important that the specifications be written in such a manner that meeting them will lead to a satisfied customer.

One way to develop specifications that will satisfy the customer, whether customer is defined as the highway agency or the driving public, is to identify the properties that are desired in the final product. In this sense, the term "property" might be thought of as a generic attribute such as strong, durable, or smooth. These properties must be translated into some measurable characteristic that can then be specified and tested to determine conformance. The TRB glossary (2) defines "quality characteristic" as follows:

- ***Quality characteristic****—That characteristic of a unit or product that is actually measured to determine conformance with a given requirement. When the quality characteristic is measured for acceptance purposes, it is an* ***acceptance quality characteristic*** *(AQC).*

Under the above definition, there is a difference between a "property" and a "quality characteristic." There is also a difference between a "quality characteristic" and a "test method." For example, for hot-mix asphalt concrete (HMAC) pavement, the desired "property" might be durability. A "quality characteristic" related to durability might be "asphalt content." The "test method" to obtain this quality characteristic measurement might be the ignition oven. Throughout this manual the term "quality characteristic" is used whenever reference is made to a value that is measured for either quality (process) control purposes or to assess product acceptability. However, in practice, the term property is often used interchangeably with quality characteristic.

Definitions Relating to Specifications

Traditionally, highway specifications spelled out in detail the work that was to be done by the contractor. These types of specifications are described in the following TRB glossary (2) definition:

- ***Materials and methods specifications****—Also called* ***method specifications****,* ***recipe specifications****, or* ***prescriptive specifications****. Specifications that direct the contractor to use specified materials in definite proportions and specific types of equipment and methods to place the material. Each step is directed by a representative of the highway agency. [Experience has shown this tends to obligate the agency to accept the completed work regardless of quality.]*

The results from the AASTHO Road Test helped to begin a trend towards more testing to evaluate highway materials and construction. Their high personnel requirements, along with the shrinking personnel resources of many highway agencies, also contributed to a move away from materials and methods specifications in favor of more testing of the final product. The TRB glossary (2) includes the following definition:

- ***End result specifications****—Specifications that require the contractor to take the entire responsibility for supplying a product or an item of construction. The highway agency's responsibility is to either accept or reject the final product or to apply a price adjustment commensurate with the degree of compliance with the specifications. [End result specifications have the advantage of affording the contractor flexibility in exercising options for new materials, techniques, and procedures to improve the quality and/or economy of the end product.]*

In practice, current specifications are neither solely "materials and methods" nor "end result." The TRB glossary (2) defines "quality assurance specifications" as follows:

- ***Quality assurance specifications****—Also called* ***QA/QC specifications*** *or* ***QC/QA specifications****. A combination of end result specifications and materials and methods specifications. The contractor is responsible for QC (process control), and the highway agency is responsible for acceptance of the product. [QA specifications typically are statistically based specifications that use methods such as random sampling and lot–by–lot testing, which let the contractor know if the operations are producing an acceptable product.]*

From the above definition it can be seen that a QA specification consists of two separate functions—quality control or process control, and acceptance. Both of these major functions are addressed in detail in subsequent chapters. The TRB glossary (2) contains the following definitions:

- ***Quality control (QC)****—Also called* **process control***. Those QA actions and considerations necessary to assess and adjust production and construction processes so as to control the level of quality being produced in the end product.*

- ***Acceptance****—Sampling and testing, or inspection, to determine the degree of compliance with contract requirements.*

- ***Quality assurance (QA)****—All those planned and systematic actions necessary to provide confidence that a product or facility will perform satisfactorily in service. [QA addresses the overall problem of obtaining the quality of a service, product, or facility in the most efficient, economical, and satisfactory manner possible. Within this broad context, QA involves continued evaluation of the activities of planning, design, development of plans and specifications, advertising and awarding of contracts, construction, and maintenance, and the interactions of these activities.]*

By these definitions, therefore, QA is a combination of QC and acceptance.

These definitions also indicate the fundamental separation of QC and acceptance. QC should be used for control of the process. Acceptance should be used to assess the quality of the product and, when appropriate, establish payment. These functions should not be determined by whether the contractor or the agency performs the tests, but instead by the purpose of the test. Thus, in a QA program, the contractor is responsible for QC and, as will be discussed, may also be responsible for performing acceptance tests. If the contractor is responsible for both testing functions, they should continue to be separated. One way, but not the only way, that has been used to make this separation more distinct, in application, is to use acceptance functions, not QC functions, in the determination of payment. However, some acceptance procedures may employ screening tests (see chapter 6) that may be used for pass/fail decisions rather than for payment determination.

It should be pointed out that the definitions for QC and acceptance contained in FHWA *23 CFR Part 637B* (6) differ from those in the TRB glossary. (2) The FHWA *23 CFR 637B* (6) definition for **quality control** is *"All contractor/vendor operational techniques and activities that are performed or conducted to fulfill the contract requirements*." This definition was adapted from American National Standards Institute (ANSI) standard ANSI 90 and International Organization for Standardization (ISO) standard ISO 9000. **Acceptance program** is defined as "*All factors that comprise the State highway agency's (SHA) determination of the quality of the product as specified in the contract requirements."* The definition for QA is similar to that in the TRB glossary. (2)

The goal of a QA specification is to relate the measured quality characteristics to the anticipated performance of the materials or construction. The TRB glossary (2) defines performance–related specifications as follows:

- ***Performance–related specifications****—QA specifications that describe the desired levels of key materials and construction quality characteristics that have been found to* *correlate with fundamental engineering properties that predict performance. These characteristics (for example, air voids in AC and compressive strength of portland cement concrete [PCC]) are amenable to acceptance testing at the time of construction. [True performance–related specifications not only describe the desired levels of these quality characteristics, but also employ the quantified relationships containing the characteristics to predict as–constructed pavement performance. They thus provide the basis for rational acceptance/pay adjustment decisions.]*

The TRB glossary (2) also includes the following definition:

- ***Acceptance plan***—*an agreed–upon method of taking samples and making measurements or observations on these samples for the purpose of evaluating the acceptability of a lot of material or construction.*

Based on the above definitions, the term "acceptance plan" could be considered to represent only those functions associated with acceptance, with QC thought of as a separate function. However, a broader interpretation of the term might include QC activities as well as acceptance testing since they could both be viewed as part of the process for "evaluating the lot of material or construction." In this manual, the term acceptance plan is used in the narrower sense of relating to only the acceptance decision. The term "QA specification" is used to represent the combined acceptance and QC procedures, as well as other features such as verification and dispute resolution procedures.

The TRB glossary (2) goes on to define two general types of acceptance plans:

- ***Attributes acceptance plan***—*a statistical acceptance procedure where the acceptability of a lot of material or construction is evaluated by (1) noting the presence or absence of some characteristic or attribute in each of the units or samples in the group under consideration and (2) counting how many units do or do not possess this characteristic.*

- ***Variables acceptance plan***—*a statistical acceptance procedure where quality is evaluated by (1) measuring the numerical magnitude of a quality characteristic for each of the units or samples in the group under consideration and (2) computing statistics such as the average and standard deviation of the group.*

The acceptance plans in this manual are based, for the most part, on the variables approach.

The acceptance plan may be a special provision, or part of a supplemental or standard specification. Typically, a new acceptance plan is first employed as a special provision before becoming a part of a supplemental specification and, finally, becoming part of a standard specification.

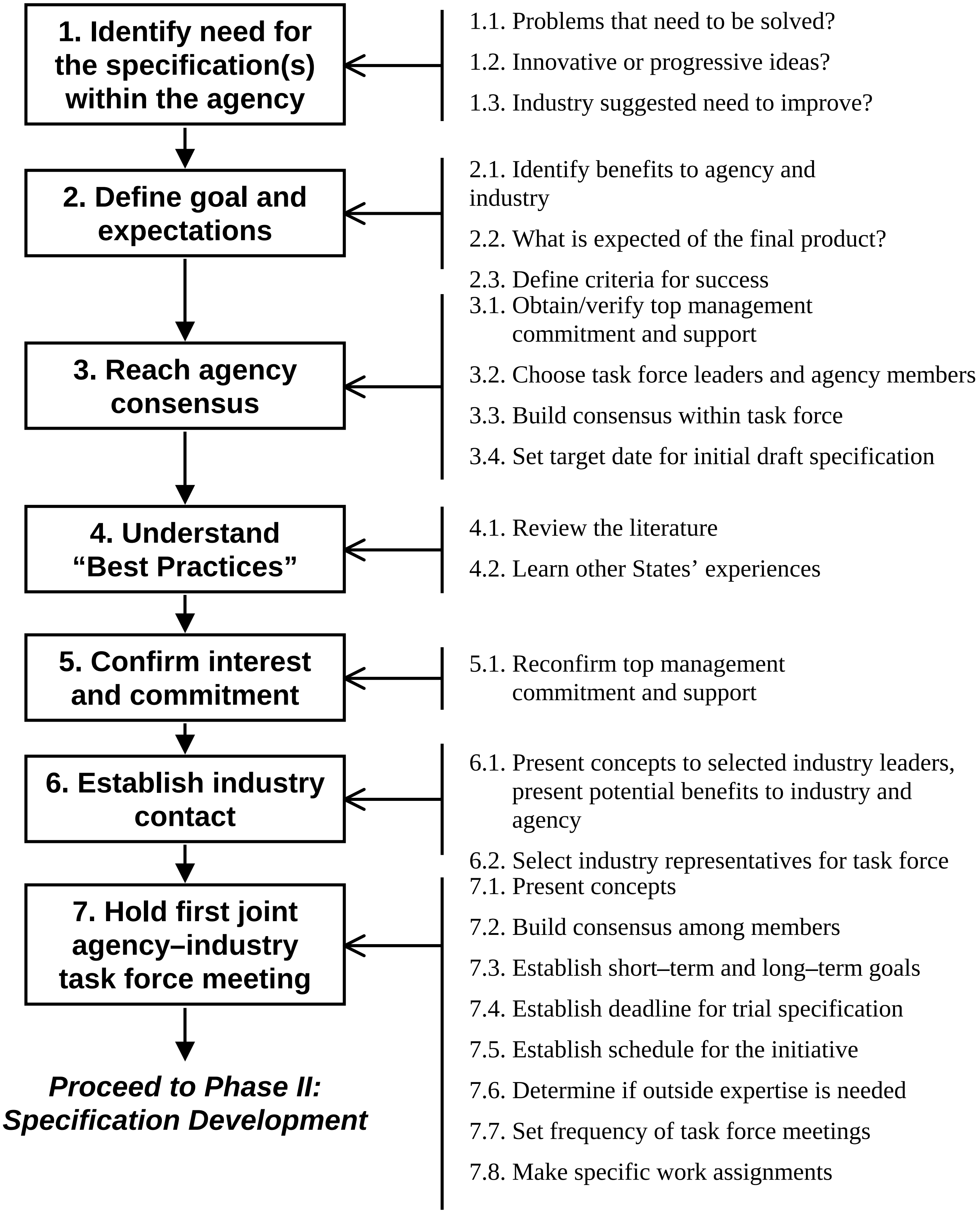
A primary intent of this manual is, through the use of proper, well–established analytical tools, to demonstrate how effective QC procedures and acceptance plans can be incorporated into QA specifications that are fair to both the contractor and to the agency. The examples used in this manual are directed to those most appropriate for control and acceptance of HMAC and PCC.

Specification Development Process

The overall specification development and implementation process can be divided into three primary phases:

- **Phase I:** Initiation and Planning.
- **Phase II:** Specification Development.
- **Phase III:** Implementation.

The steps in each of these phases can be represented in the form of a flowchart for each phase. This manual presents and discusses the steps in each of the three phases of the overall specification development and implementation process. Flowcharts for the overall specification development and implementation process are shown in figures 1 through 3. Specific sections of these flowcharts are presented again to support the detailed discussions of each phase that are presented in upcoming chapters. Phase I is addressed in chapter 2. Phase II is covered in chapters 3 through 7. Phase III is presented in chapter 8. Chapter 9 presents some case studies of the implementation of QA specifications.



1. Identify need for the specification(s) within the agency

- 1.1. Problems that need to be solved?
- 1.2. Innovative or progressive ideas?
- 1.3. Industry suggested need to improve?

2. Define goal and expectations

- 2.1. Identify benefits to agency and industry
- 2.2. What is expected of the final product?
- 2.3. Define criteria for success

3. Reach agency consensus

- 3.1. Obtain/verify top management commitment and support
- 3.2. Choose task force leaders and agency members
- 3.3. Build consensus within task force
- 3.4. Set target date for initial draft specification

4. Understand "Best Practices"

- 4.1. Review the literature
- 4.2. Learn other States' experiences

5. Confirm interest and commitment

- 5.1. Reconfirm top management commitment and support

6. Establish industry contact

- 6.1. Present concepts to selected industry leaders, present potential benefits to industry and agency
- 6.2. Select industry representatives for task force

7. Hold first joint agency–industry task force meeting

- 7.1. Present concepts
- 7.2. Build consensus among members
- 7.3. Establish short–term and long–term goals
- 7.4. Establish deadline for trial specification
- 7.5. Establish schedule for the initiative
- 7.6. Determine if outside expertise is needed
- 7.7. Set frequency of task force meetings
- 7.8. Make specific work assignments

Proceed to Phase II: Specification Development

Figure 1. Flowchart for Phase I—Initiation and Planning

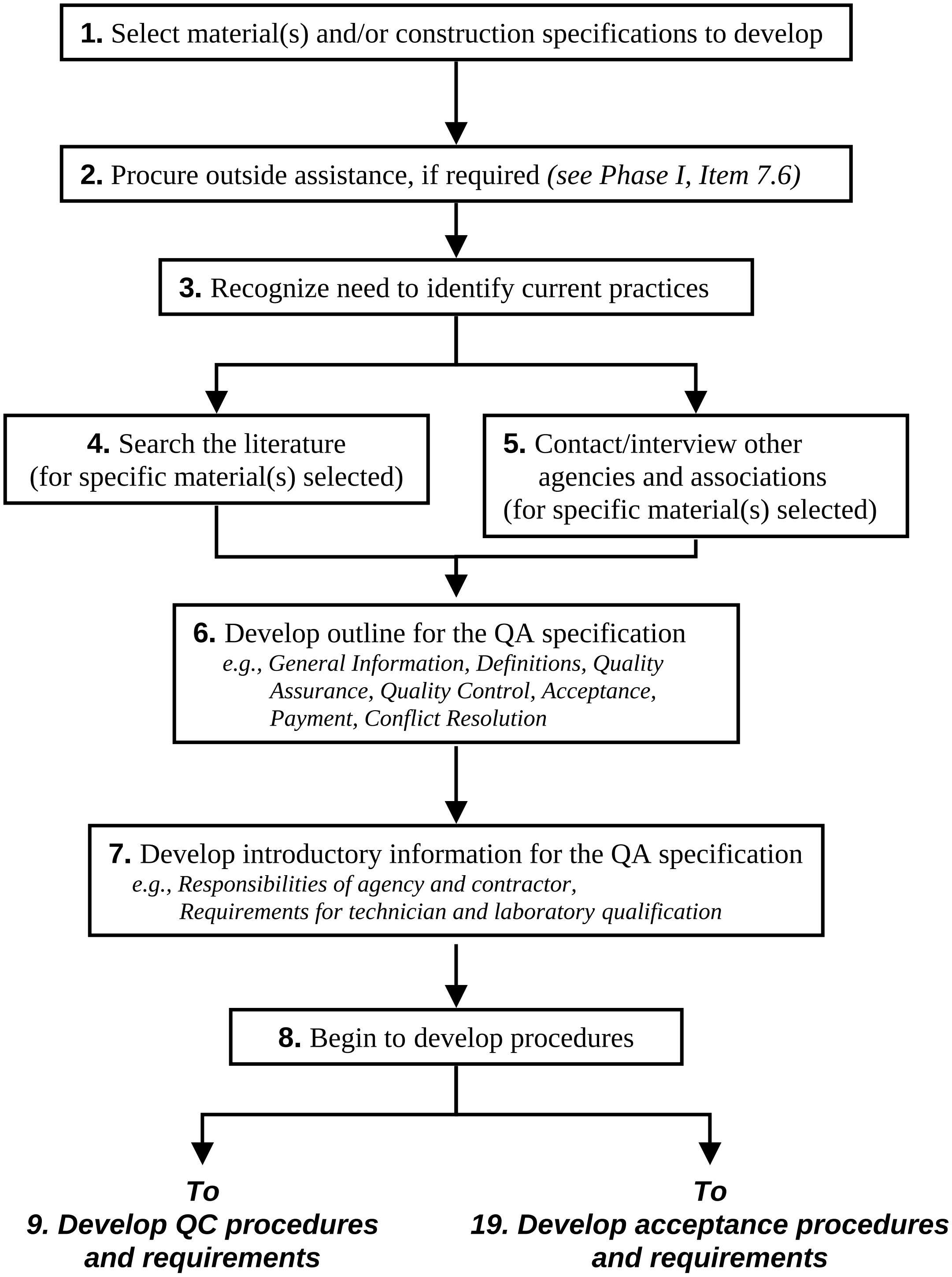


Figure 2. Flowchart for Phase II—Specification Development

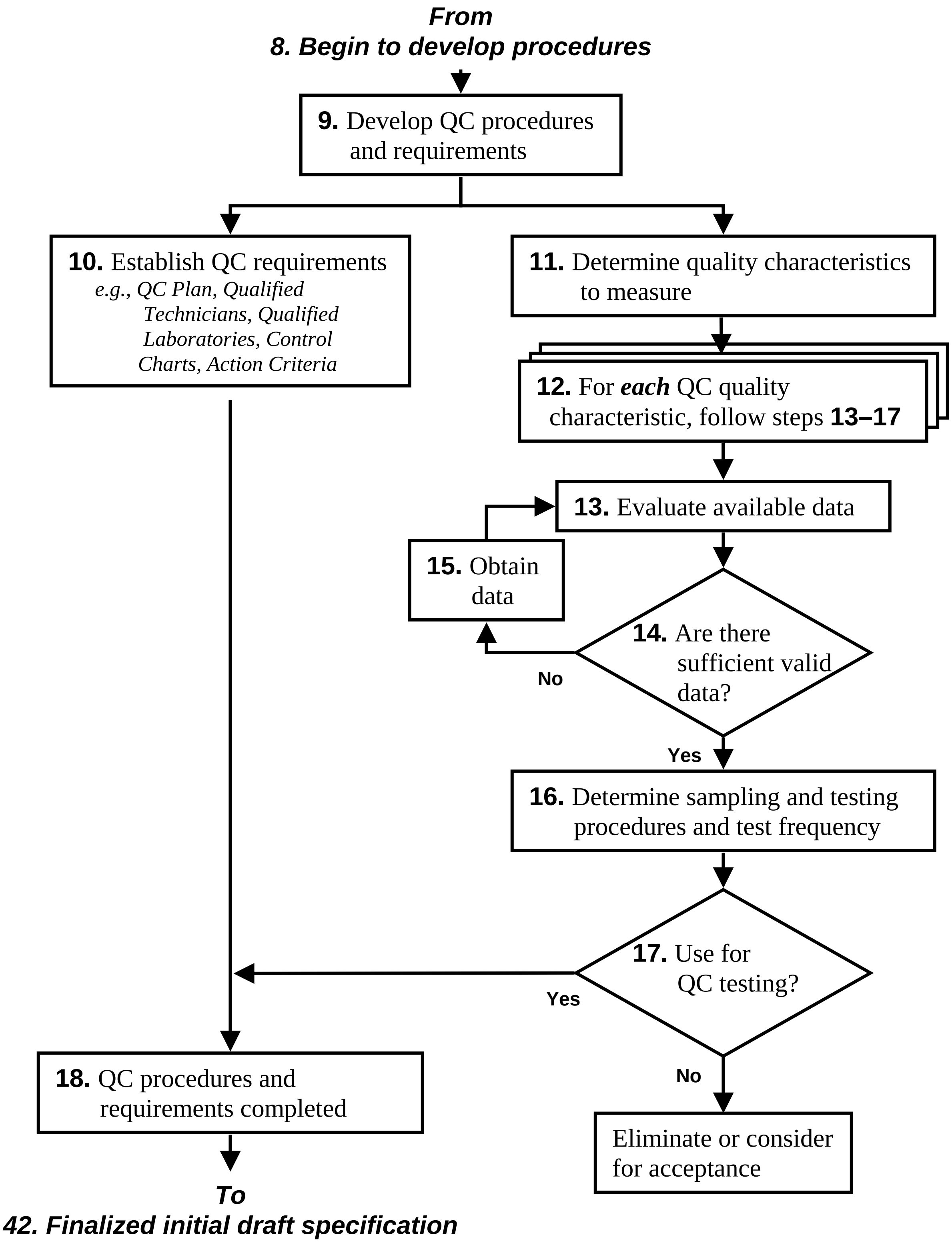


Figure 2. Flowchart for Phase II—Specification Development *(continued)*

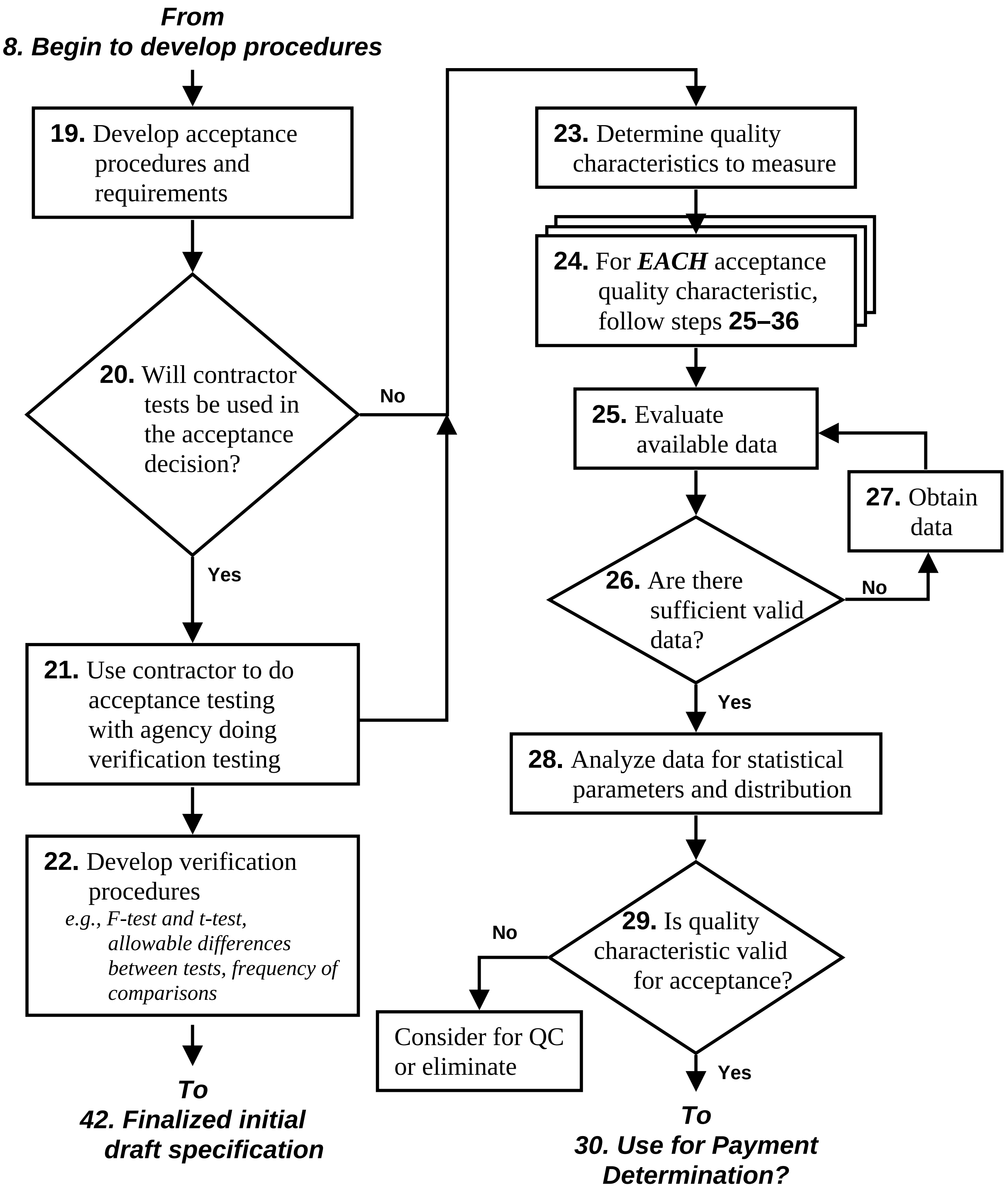


Figure 2. Flowchart for Phase II—Specification Development ***(continued)***

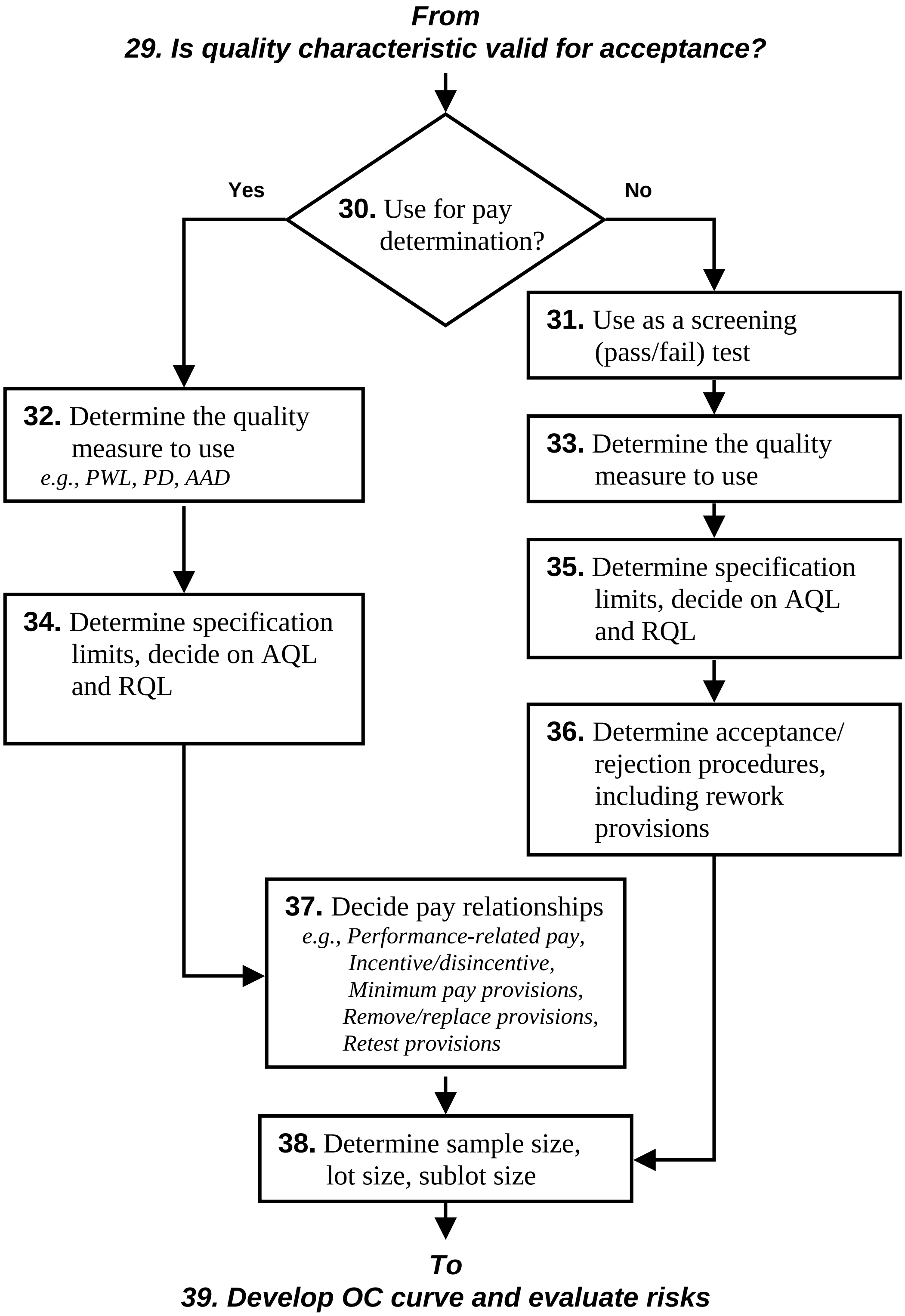


Figure 2. Flowchart for Phase II—Specification Development *(continued)*

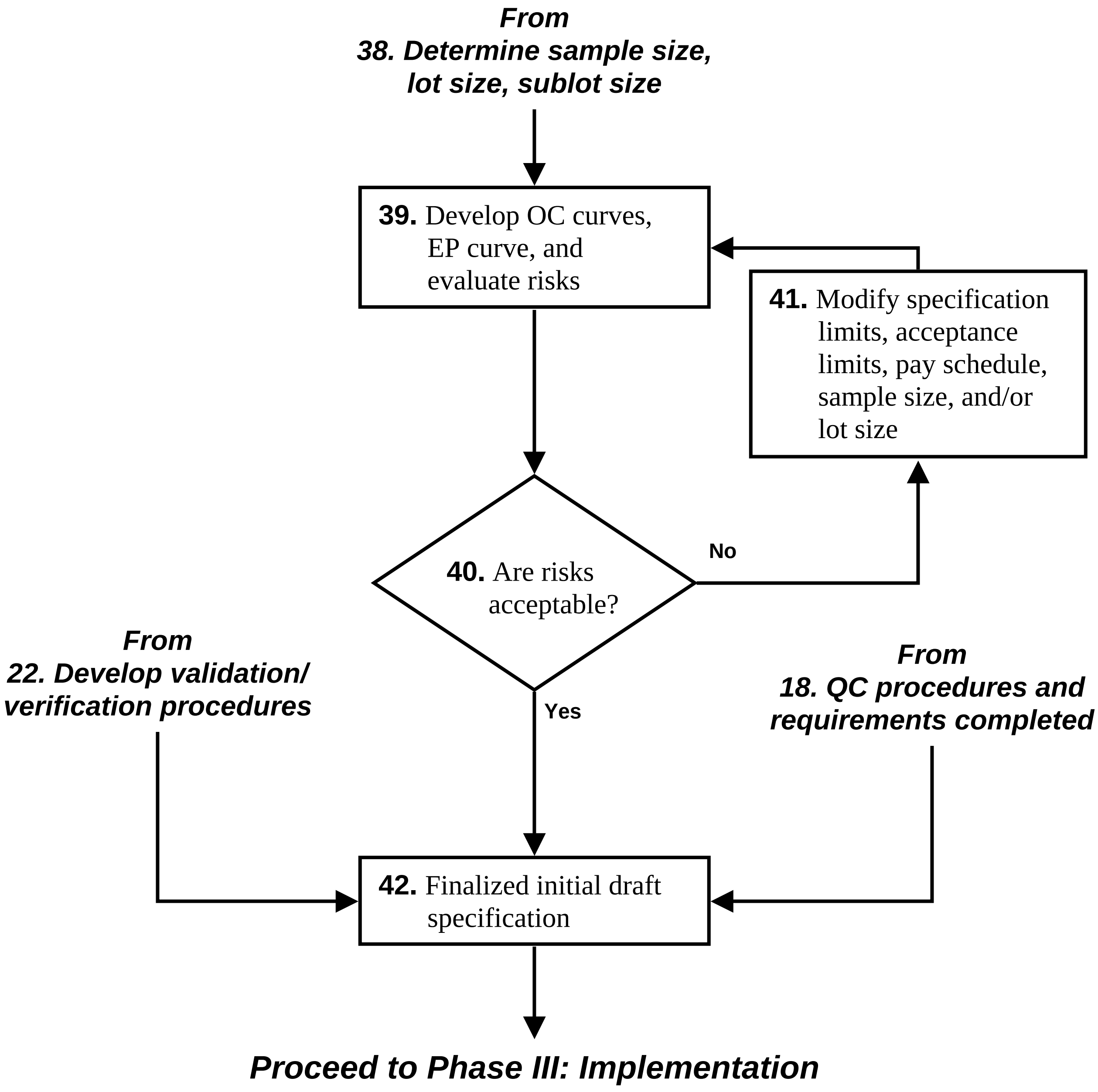


Figure 2. Flowchart for Phase II—Specification Development ***(continued)***

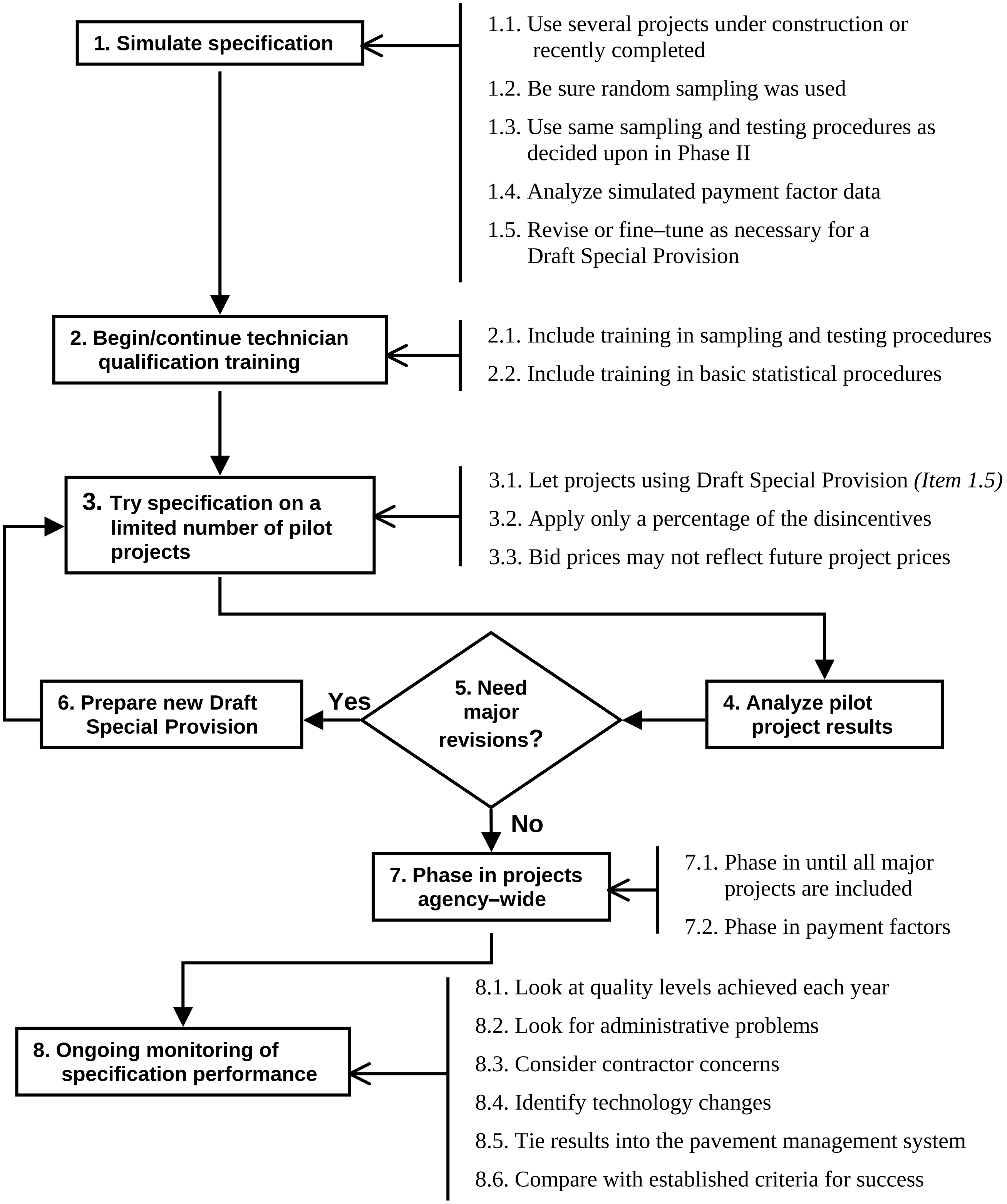


Figure 3. Flowchart for Phase III—Implementation

CHAPTER 2. INITIATION AND PLANNING

Initiation and planning form Phase I in the specification development and implementation process. The purpose of this chapter is to explain, in detail, the need for planning and the issues that should be addressed before embarking on the development of a new QA specification or the revision of an existing QA specification. The steps that are involved in this process are identified in the flowchart in figure 4. Each of the seven major steps in the flowchart is discussed in the following sections. The numbers in boxes before section titles refer to the corresponding box in the flowchart.

1 Identify the Need

One of the first considerations for a new, or modified, QA specification is to establish the need for change. Very seldom is the development of a new specification undertaken before it is needed, but caution should be exercised to assure that this developmental process is timely. Because change, in any form, is often hard to accomplish, a definable need should be established before embarking on a change. Thus, it is essential to recognize, identify, and document the need for the new or revised QA specification.

1.1. Reasons for Developing or Modifying a QA Specification

There are a number of potential reasons for needing a new or revised QA specification. There may be problems that have been identified with the present specification that need to be solved. QA specifications should be dynamic and evolve as other technology evolves. Technology in testing or the process of making the product may have evolved to a point that the present procedures are obsolete. The present procedures may have become outdated and a more innovative and progressive QA specification is needed. There may be better ways to obtain the desired product than those presently used. For instance, QA specifications have become a very popular way of sharing responsibility between the agency obtaining and the contractor providing quality products.

Some of the reasons that a new QA specification may be needed include:

- The present quality levels are substandard.
- Premature failures have occurred that have been related to the current specification procedures.
- It has been decided that a quality measure is needed.
- It has been decided that a different quality measure should be used.
- A new quality characteristic upon which to base acceptance has been identified.

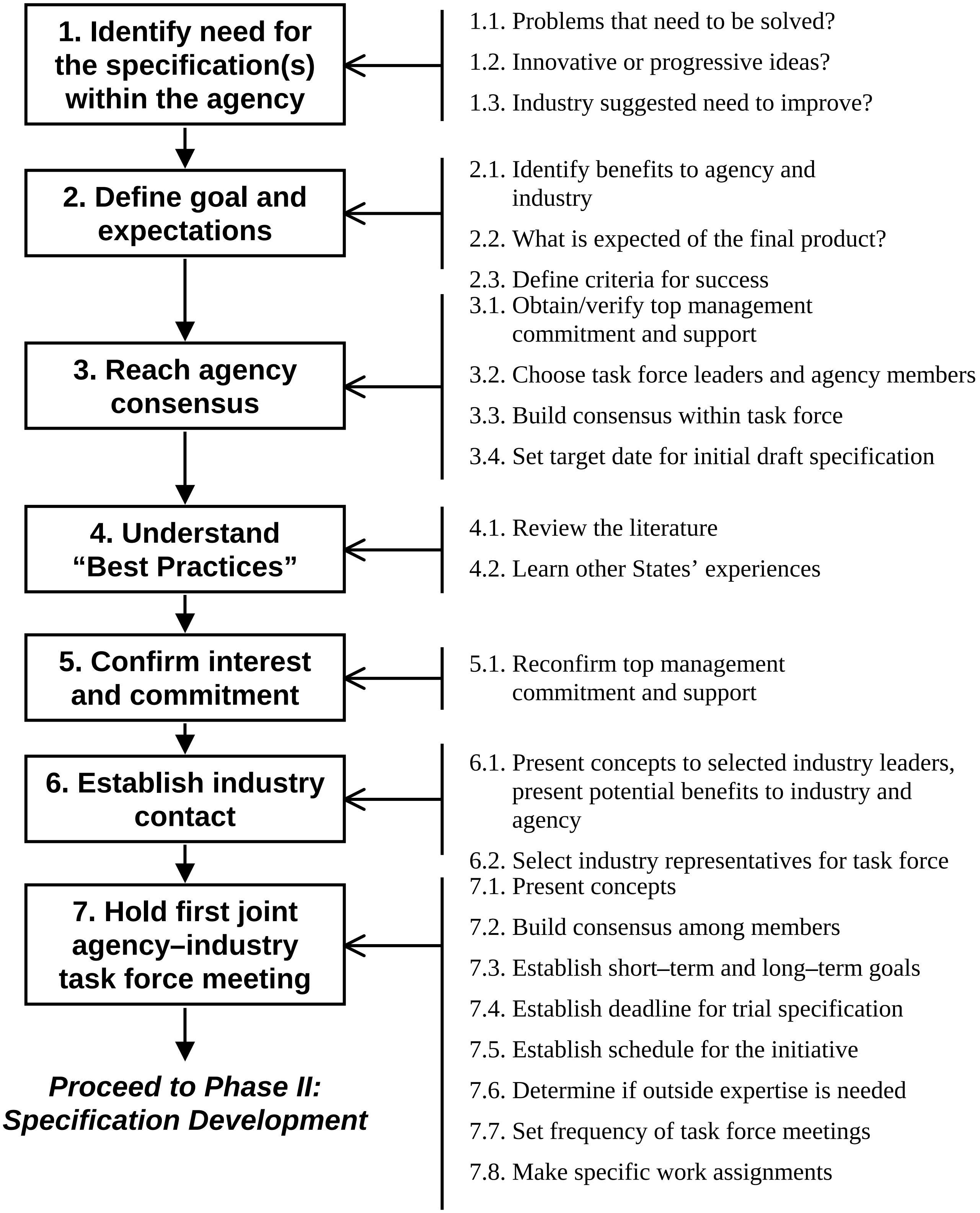


Figure 4. Flowchart for Phase I—Initiation and Planning

Potential reasons for modifying an existing QA specification include:

- The existing specification has failed to produce a consistently high quality.
- The frequency or magnitude of payment adjustments does not appear to be commensurate with quality.
- The rejection frequency is not appropriate.
- Sampling rates are not appropriate.
- Lot sizes are not appropriate.
- The risks are not properly balanced, or are too high.
- The acceptable quality level (AQL) and/or rejectable quality level (RQL) value(s) appear unrealistic.
- The specification limits or acceptance limits appear unrealistic.
- Better test methods have become available.
- Contractor capabilities have improved.
- It is desired that the acceptance procedure incorporate a measure of variability.
- More efficient or effective statistical methods are desired.
- The acceptance procedure should be simplified.
- The acceptance procedure should be made more rigorous.
- The current specification has administrative or legal problems.

The reasons for developing or modifying the QA specification will likely address whether the problem(s) is (are) with a single procedure or whether several procedures do not seem to be working properly, such as not being effective in obtaining the desired product.

1.2. Identify Source of the Initiative

Another consideration is to identify the source of the initiative to develop or modify the QA specification. Did the initiative come from within the agency or did it come from the private industry sector? This is important from the standpoint of getting input into the planning process. If, for instance, a segment of the private industry feels strongly that the specification needs changing, it is logical to have the reasons explained and hear what changes are proposed. If the initiative for change comes from within the agency, the reasons for the change should be stipulated by the agency.

However, from whichever source the desire for a change may come, expect some resistance or skepticism from individuals in the other source. The low bid system has generated a degree of skepticism between agencies and contractors when it comes to change. Although this is a generalization, it has been observed to exist nationwide to various degrees. The potential for this skepticism should be used in a positive manner to assure that the reasons for the change are explored, validated, and explained in detail. Also, when it comes to the implementation phase, joint training of agency and contractor personnel can help to minimize this problem.

2 Define Goals and Expectations

The first goal is to identify potential benefits to the agency and to the industry. One of the goals will certainly be to correct the identified deficiencies with the present specification(s). If the present specification is too restrictive, making it less restrictive will be a goal. If the present specification does not place any responsibility on the contractor and a sharing of responsibility is desired, this becomes a goal. However, it must be recognized that with responsibility goes the authority and freedom of operation associated with the responsibility. Whether the industry is in a position to assume more responsibility should be determined. If not, what will be required to facilitate their moving into such a position should be determined. Under any potential goals, the benefits to the agency and to the contractor must be identified. If only one party receives benefits, the other party may wonder why a change is desirable. Once again, in defining the goals, the scope of the initiative must be considered. If more than one area of materials and construction is to be considered, the general goals may be similar but the specific goals may differ for each area considered. Both the general and specific goals should be identified.

2.1. What is Expected of the QA Specification?

Identify and list the expectations for the new QA specification. Ensure that they are realistic and not just “wishful thinking.” For instance, expecting that the new specification will double the expected pavement life or reduce the life–cycle cost (LCC) by 50 percent may seem desirable, but is unlikely to occur. If unrealistic expectations are listed, the final product may be considered a failure when the expectations are not met; yet, the product may still be substantially better than that previously produced.

Define the criteria that will be used to measure the degree of success. Decide upon the criteria that will be used to judge whether the goals and expectations have been met when the QA specification is implemented. The criteria must be realistic. A few potential criteria are listed below, and some are more realistic than others. Expecting all problems to be solved by the improved specification is an unrealistic goal.

Possible criteria upon which to judge success include:

- **Improved quality.** Certainly improved quality of the product is always a desirable goal. However, to make this determination, quantifiable measures of quality are needed for both the old and new specifications. Such quantifiable measures may not be present in some old specifications, but are a necessary part of a QA specification.
- **More knowledgeable industry.** One advantage of QA specifications is that they place more responsibility on the contractor; thereby making the industry more knowledgeable

about the product it produces. This increased knowledge may be difficult to quantify, but has been observed and is important.

- **Faster completion times.** It may be possible to speed production/construction by the use of less restrictive specifications.
- **Lower bid cost.** This may be desirable, but because of many other considerations in the bidding process a reduction in bid price may not be realistic or even determinable.

3

Reach Agency Consensus

The first, and most essential, step in reaching agency consensus is to obtain top management support. If there is a single success issue that stands out as more important than any other, it is obtaining firm top management commitment and support at an early stage of the initiative. In the case that top management is not promoting the initiative, it must be verified that the direction of the initiative has this vital support. More quality initiatives have been curtailed, delayed, or had the direction changed because of a decision by top management than for any other single reason. Keep in mind that top management has many issues with which to deal and that these other issues may impact the quality initiative.

Use the benefits to both the agency and the industry to gather top management support. Benefit/cost advantages are an important “selling tool” for top management. This information may be difficult to obtain, but is highly regarded by top management.

Realize that many obstacles may stand in the way of successful completion of the initiative. Although resistance to change is natural, with proper management support, enthusiastic leadership, enlightenment, and persistence, the obstacles can be overcome.

3.1. Choose Progressive Leadership

Next, choose agency leaders and members to be on a task force that eventually will be a joint agency/industry task force. The successful accomplishment of this step cannot be overemphasized. Leaders at all levels must realize the benefits that can emerge from the initiative. Ensure that representatives from all operating divisions and field personnel from different levels that will use the QA specification are included on the task force. At the same time, avoid having such a large task force that it becomes unwieldy. One way around this dilemma is to eventually establish subgroups within the task force to address special issues. Also, when selecting personnel for the task force, without being biased, choose members with a progressive “can do” attitude. It may not hurt to have a few skeptics on the task force. Convincing skeptics of the advantages of QA specifications can be a strong selling point when it comes to implementation. But try to avoid obstructionists who may hinder progress.

Build a consensus within the agency task force members to pursue the development and eventual implementation of the QA specification. Have a well–thought–out outline of scope for the specification before convening the task force. This will provide initial direction and give the task force specific issues to address and discuss. Expect differences of opinion. Try to create an open, free–flowing discussion of the main issues. Try not to get bogged down with minute details this early in the discussion—this may just delay progress. These issues should be addressed during the developmental stage. But, it is important that the members "buy into" the need to develop the specification, and to stress that members need to maintain an open mind to the potential for success.

3.2. Set Initial Target Date

Set a target date for the initial draft of the QA specification. The amount of time it will take to develop the QA specification will depend on many factors. One important factor is the experience that the agency has with QA initiatives. If this is the first attempt to develop a QA specification, the developmental period may require several years. If an existing QA specification is merely being modified, the period may be less than a year. The timing is important and strong consideration should be given to doing it right as opposed to doing it quickly. Do not rush the target date. Make sure that both agency and contractor personnel will have time to understand the specification before rushing to implement it. The dates, tentatively established at this point, may need to be revised after meeting jointly with representatives from the private sector.

4

Understand "Best Practices"

There are two tasks that should be addressed at this point. Both of these tasks are preliminary and will be greatly expanded once the specification development phase begins.

4.1. Conduct Literature Review

The first task is to conduct a preliminary literature review. Get an idea what other agencies have done. This should not be exhaustive at this point but should give guidance as to direction. For instance, what is the general state–of–the–practice of acceptance plans? What are some of the important issues that need to be addressed in a well written QA specification? NCHRP and other syntheses, as well as published reports, are a good starting place for this information.

4.2. Learn from Other Agencies

The second task goes hand–in–hand with the literature search. This is to learn from the experiences of other agencies, especially those of neighboring States. Concerns and issues in one State may be similar to those of adjacent States. This is not to say that the experience of other States that may be more geographically removed should be excluded, because these agencies may have found solutions to problems that are mutual. Contact these agencies and request information that will provide guidance as to which steps should be taken and which should be avoided when planning the development of the QA specification. Whenever possible, talk to the architect of the specification. Maintain a list of contacts and keep thorough notes of the discussions. Keep this information in a file so that the experiences of these sources can be considered during the specification development and implementation phases. (These issues are discussed in more detail in chapter 3.)

5 Confirm Top Management Interest and Commitment

Once the current best practices have been more fully determined, e.g., the type of the initial QA program has been defined, or the need to improve the QA program through a more equitable payment adjustment plan has been identified, it is prudent to explain this more complete initiative to top management. This then gives top management the opportunity to reconfirm their commitment and support now that they have a fuller understanding of what will be involved in the development and implementation phases. As a generalization, it is important to reconfirm this support at frequent intervals during the development and implementation process to assure that this commitment is maintained throughout the initiative. In addition, top management should periodically be briefed on the progress of the initiative. As previously mentioned, issues facing top management change and these changes may impact the progress of the quality initiative.

6 Establish Industry Contact

It is now time to seek industry acceptance of and participation in the new quality initiative. It is important that the agency has reached consensus on the need for and the benefits of the new specification before presenting it to industry. The basic concepts as well as the potential benefits to both the industry and the agency should first be presented to selected industry leaders. If agency/industry standing committees already exist, they are a good place to start initial industry contact. If such committees do not currently exist, representatives from associations, such as the Associated General Contractors (AGC), paving associations, and aggregate associations, can be used as initial industry task force members. The information presented at this time is conceptual and preliminary. It is presented for the purpose of identifying industry members to serve on the joint task force. It is important to recognize at this point that it is not up to the industry to develop the specification, but that their input is invaluable in the development phase and, particularly, in the implementation phase of the QA initiative. Use the previously discussed benefits to both the agency and the industry to gather support from the industry. Emphasize the benefits that can accrue to both parties to the contract.

6.1. Choose Progressive Industry Leadership

When selecting industry representatives for the joint agency/industry task force, members of existing standing committees may be used for the task force, but be careful to ensure that both management and field personnel from industry who will use the specification are included on the task force. Once again, try to choose members with progressive attitudes, but take care that the task force is not too large.

7

Hold the First Joint Agency/Industry Task Force Meeting

The first joint agency/industry task force meeting should present basic concepts. It should provide the background on the QA initiative and on the specification to be developed. The members should discuss why the new specification is being proposed and what the expected outcome will be.

7.1. Build Consensus Among Task Force Members

It is important to build a consensus for the new specification among the task force members. To do so, reiterate the potential benefits both to the agency and to the industry. Try to generate discussion within the joint task force regarding how the members feel about potential for success, their concerns, potential obstacles or pitfalls, etc.

Review the target date for the initial draft of the QA specification and, if necessary, revise this date with industry input. Determine a firm target date with which everyone on the task force is reasonably comfortable. Assure that there is not a feeling that the process is being rushed.

With the date of the initial draft specification set, establish the schedule for the entire initiative including the development and implementation phases. These dates may be tentative, but this step provides an initial expectation as to when the QA specification will be ready for review, trial, and implementation.

7.2. Establish Short–Term and Long–Term Goals

Establish both short–term and long–term goals and the steps needed to accomplish these goals. Potential goals, which are presented and discussed in detail in chapters 3–7, include:

- **Develop initial draft QA specification.** As seen in chapter 3, this is quite an involved task. There are many steps that must be taken and data that need to be gathered. It is important that this task be undertaken in a systematic, rational manner.
- **Simulate the specification.** Once the initial draft specification has been developed, reviewed and approved, the first steps of implementation can begin.
- **Train and qualify personnel.** Personnel of both the agency and the industry need to become familiar with the workings of the specification. This is usually a time–consuming step, and one that must be continued on an ongoing basis. It is important to allow sufficient time to do this properly.
- **Conduct pilot projects.** Pilot projects are a good way to "field–test" the specification and to work out initial "bugs" that inevitably seem to develop.
- **Revise the specification.** Fine–tune the specification based on results from the pilot projects.

- **Phase–in payment factors.** This is a consideration that has been demonstrated as a means for alleviating concerns of contractors as to how their operations will be impacted by the payment adjustment system.
- **Full implementation.** This is the culmination of the QA specification development.

7.3. Determine if Outside Expertise Is Needed

The task force must determine whether any outside expertise is needed for any of the steps to reach the short–term or long–term goals. Consider whether additional expertise outside the agency and industry is needed for the task force and/or for developing the QA specification. Technical expertise dealing with new technology, materials, construction, specification development, and/or statistics, may be needed. Consider academia as one source of help.

7.4. Set Frequency of Task Force Meetings

The frequency for holding task force meetings must be established. Generally, the meetings should coincide with milestones that will have reached specific goals and, therefore, will contain details upon which progress can be measured or decisions made. Do not meet “just to meet.”

7.5. Make Assignments within the Task Force

Make specific assignments for the task force members. Make certain that every step has someone assigned to work on it, keep it on schedule, and report the progress. Develop a procedure to monitor progress to assure the schedule is maintained or to determine the reason that progress is behind schedule.

CHAPTER 3. SPECIFICATION DEVELOPMENT

Specification development comprises Phase II in the specification development and implementation process (see chapter 1). This chapter is the first of five that describe, in detail, the specification development process. It is intended to provide “how to use” best practices in the early specification development phase. The steps that are involved in this part of the process are identified in the flowchart in figure 5. The numbers in boxes before the titles of the following sections refer to the corresponding box in the flowchart.

1 Select the Material and/or Construction Area(s)

First, select one or more areas of material and/or construction with which to begin the quality initiative. Most, but not all, agencies select one area of materials or construction with which to begin. The reason that most agencies take this approach is that it is usually easier to select a single area for the initial effort. However, if both HMAC and PCC specifications are desired, it is conceivable that both can be undertaken simultaneously. If it is the agency’s first effort at developing a QA specification, then selecting a single material or construction area is recommended.

1.1. Advantage of Selecting One Area

The advantage of selecting one area is that the area that appears to be the easiest or best suited for QA specification development can be selected. The process that is employed, and the lessons learned, on this first effort can then serve as a model when additional areas are selected. Traditionally, agencies have developed HMAC specifications before PCC specifications. (3) It is unlikely that the initial QA specification area will be developed without some misdirection or erroneous assumptions during the developmental and implementation processes. These misdirections can be invaluable learning tools that indicate what to avoid when starting a new area.

If both HMAC and PCC QA specifications are undertaken simultaneously, the time necessary to have both ready for implementation may be reduced over the approach in which one is undertaken and completed before starting on the other.

2 Procure Outside Expertise If Required

Determine if outside expertise is required. There likely will be expertise and experience within the agency for the material/construction area of interest, but there may or may not be expertise in specific areas of the QA initiative. For instance, defining appropriate lot and sample sizes,

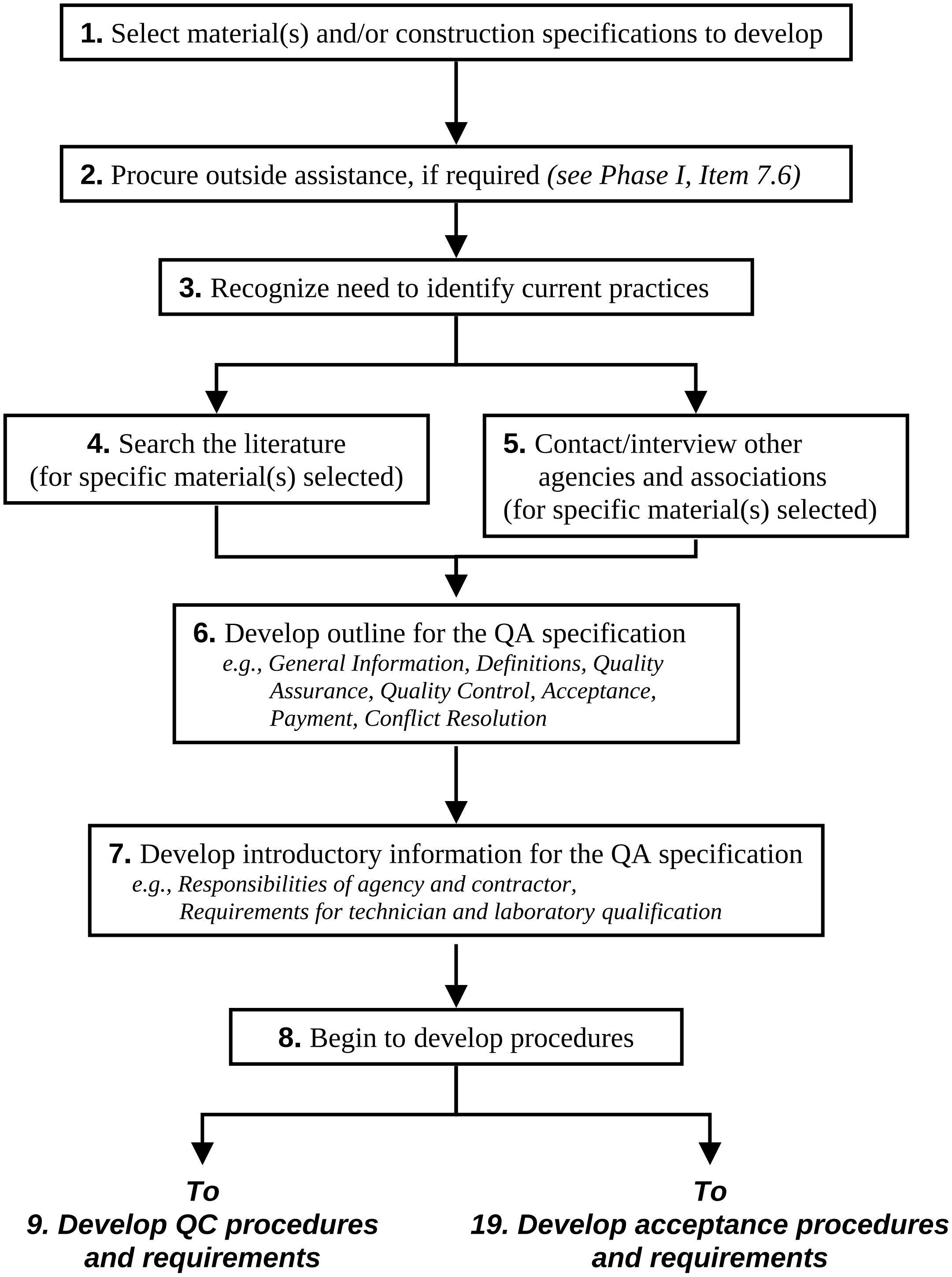


Figure 5. Flowchart for Initial Portion of Phase II

selecting appropriate statistical procedures, or developing fair payment factors may require expertise not available within the agency. Identify precisely what areas of expertise are needed. For this, first it is important to know precisely the area(s) for which the QA specification is being developed or modified, and then to determine what areas of expertise are needed when procuring outside assistance.

It is best to begin the process of obtaining outside expertise very early in the QA initiative since most agency contracts for outside expertise take quite a while to execute. Beginning the process and obtaining the necessary help early will aid in accomplishing the subsequent steps and help to keep the developmental process on schedule.

3

Recognize the Need to Identify Current Practices

There is no reason to spend needless time "reinventing the wheel." Look for other agencies or sources that have developed practices that may provide guidance regarding how to proceed or what type of QA specification may be most appropriate. It is likely that other agencies or sources have experience in developing these practices. One of the first steps in this area is to establish and maintain a list of contacts with other agencies and national associations that may be helpful in providing this guidance.

4

Conduct a Thorough Literature Search

Look for other primary sources for identifying current practices. To do this, first organize a thorough literature search of such publications as those from the TRB, the American Association of State Highway and Transportation Officials (AASHTO), the American Society for Testing and Materials (ASTM), and FHWA. Also search such national association publications as the American Concrete Paving Association (ACPA), the National Asphalt Pavement Association (NAPA), and the National Stone Association (NSA).

Recent TRB and NCHRP publications are replete with new specification initiatives that can provide guidance as to the type of specification to develop and the advantages and disadvantages within each type. Some publications that should be reviewed are presented in appendix A.

5

Contact and/or Interview Other Agencies and Associations

Much in–depth information can be gathered from personal contacts with other agencies. Making personal contacts with agencies that have experience with the type of specification being developed can save considerable time in providing the right direction in which to proceed and in learning from the mistakes of others rather than making the same ones. From the initial personal contacts, scheduling interviews with the specification writers or a small group that was instrumental in developing the specification can provide in–depth information, attitudes, and guidance that are difficult to obtain from the literature, or any other way.

6

Develop an Outline for the QA Specification

Based on the preliminary information available, develop an outline for the QA specification. This may follow general outlines used by the agency for other specifications, but items may need to be added to include new QA concepts.

6.1. Possible Information to Include

The information that may be included will vary with the type of specification written. However, some of the headings that might appear in the outline include:

- General information or requirements.
- Definitions.
- Material requirements.
- Construction requirements.
- Quality control requirements.
- Acceptance requirements.
- Verification.
- Conflict resolution.
- Measurement and payment.

7

Develop Introductory Information for the QA Specification

The introductory information necessary for the QA specification also may follow the format used by the agency in other specifications. However, the required information may differ considerably from that previously used. One of the fundamental concepts in QA specifications is the separation of the functions of QC and acceptance. This may be an entirely new concept. In QA specifications, the contractor is responsible for QC and the agency is responsible for acceptance. In some cases, as discussed subsequently, the agency may assign to the contractor the responsibility for obtaining and conducting acceptance tests. Whether the agency or the contractor conducts the acceptance tests, the separation of the responsibility for QC and acceptance testing is very important.

7.1. Suggested Topics

The information contained under each subject will vary depending on the type of specification being developed. However, a discussion of some of the important considerations is provided herein. The topics suggested below could be considered as the minimum that should be developed:

- **Responsibility of the agency.** The agency is responsible for determining whether the agency or the contractor will perform acceptance tests and what conditions will apply when this is done. As discussed in subsequent chapters, there are advantages and disadvantages to both procedures. It is important that the agency gives strong considerations to these pros and cons and selects the procedure best suited to the agency.
- **Responsibility of the contractor.** For QA specifications QC is <u>always</u> the responsibility of the contractor. Very few items in a QA specification are absolute, but requiring the contractor to assume responsibility for control of the production process is an underlying precept of a QA specification.
- **Requirements for technician qualification.** FHWA *23 CFR 637B* (6) defines qualified sampling and testing personnel as *"Personnel who are capable as defined by appropriate programs established by each SHA."* The intent is to allow each agency as much latitude as possible in establishing technician qualification requirements. Agencies often "certify" technicians as one means of ensuring that sampling and testing personnel are "qualified." Certification usually requires that technicians must be recertified on a periodic basis.

  The TRB glossary (2) does not provide definitions for this subject, but the *AASHTO Quality Assurance Guide Specification* (7) defines certified technicians as "*those who are certified through appropriate certification programs determined by each Agency, including those employed by qualified laboratories that perform acceptance or quality control sampling and testing for an Agency or Contractor, respectively.*"

  The underlying concept for both of the above definitions is to assure that the technicians who perform control and/or acceptance testing have been educated as to how to properly sample and test the material.

- **Requirements for laboratory qualification.** FHWA *23 CFR 637B* (6) defines qualified laboratories as "*Laboratories that are capable as defined by appropriate programs established by each SHA. As a minimum, the qualification program shall include provisions for checking test equipment and the laboratory shall keep records of calibration checks."*

  The TRB glossary (2) does not provide definitions for this subject either, but the *AASHTO Quality Assurance Guide Specification* (7) uses essentially the same definition as that in FHWA *23 CFR 637B* (6).

8 Begin to Develop Procedures

Two distinctly different sets of procedures must be developed for the different functions of QC and acceptance. As previously mentioned, the separation of these two functions is important. Due to the evolutionary nature of QA specifications, QC and acceptance functions often have been combined or intermingled. This has been a major source of confusion. The intermingling of QC and acceptance can be traced to the first statistically based specifications that were used at a time when agencies had technicians at the contractors' materials plants. The agency technicians did the testing and determined when the product was acceptable. The contractor made changes to the process when necessary based on the agency's tests.

Although QC was often known to be a separate item from acceptance, in reality little separation occurred. As time went on, many agencies removed their technicians from the contractors' materials plants. This resulted in the contractor having to conduct the QC tests. Acceptance was often based on the agency periodically visiting the contractor's materials plant and taking samples that were used for acceptance decisions. At this period of time, there was, typically, a separation of QC and acceptance. More recently, with agencies experiencing personnel shortages, some have given the contractor the responsibility for conducting acceptance tests. For Federal–aid contracts, under certain stipulated conditions, this acceptance testing responsibility can be delegated to the contractor.(6) However, regardless of whether or not the contractor performs the acceptance sampling and testing, the responsibility for making the acceptance decision still rests squarely with the agency.

CHAPTER 4. QUALITY CONTROL

This chapter continues the discussion of Phase II of the specification development process. This chapter is intended to provide “how to use” best practices in the development or modification of the QC issues of QA specifications. The steps that are involved in this part of the process are identified in the flowchart in figure 6. The numbers in boxes before the titles of the following sections refer to the corresponding box in the flowchart.

9 Develop QC Procedures and Requirements

As defined above and used in this manual, QC activities are those QA actions and considerations necessary to assess production and construction processes so as to control the level of quality of the end product. The QC procedures and requirements are made up of two parts: the QC requirements and the quality characteristics to be measured. These are the main ingredients that constitute the QC plan. <u>It is emphasized that the QC function is the responsibility of the contractor</u>.

9.1. Purpose of the QC Plan

It is important to realize that the purpose of the QC plan is to measure those quality characteristics and to inspect those activities that impact the production at a time when corrective action can be taken to prevent appreciable nonconforming material from being incorporated into the project. The QC efforts should also be able to quickly identify that nonconforming material is being made. These purposes should serve as a guide to the decisions used in establishing the requirements and determining the quality characteristics to measure.

The determination of who establishes the QC plan is an important one. Ideally, the QC plan should be the contractor’s plan, and not the agency’s. The contractor should know what activities to test and to inspect to produce acceptable material. Generally, two approaches have been used by agencies to specify the QC plan that is required. One is for the agency to stipulate the minimum QC requirements and properties that the QC plan must contain. The other is for the agency to specify all the requirements and properties that must be tested. Both have advantages and disadvantages. By stating the minimum requirements and properties, the agency lets the contractor know the least that is required. The disadvantage is that the contractor may view this as all that is necessary for adequate QC, rather than as the minimum. If a minimum–type QC plan is decided upon, the agency may want to require agency review or approval of the plan prior to construction. On the other hand, by stating all the requirements and properties required, the plan is likely to be all inclusive, but the contractor may view the QC plan as the agency’s plan rather than the contractor’s plan.

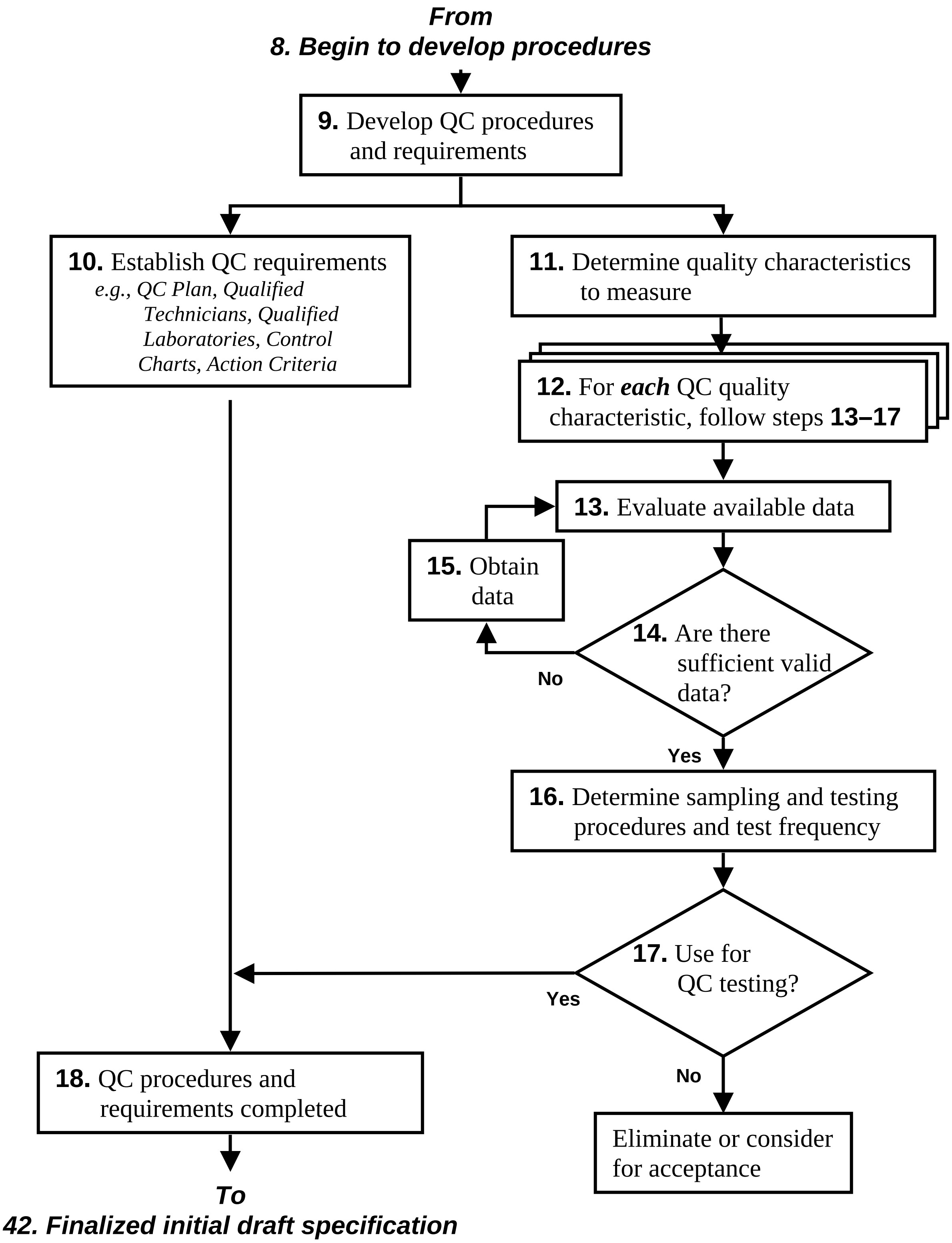


Figure 6. Flowchart for QC Portion of Phase II

If the QC plan requires agency review and/or approval, there are some legal questions that should be considered when making this decision. For instance, if the agency only reviews the QC plan, what constitutes acceptance or rejection of the plan and what are the consequences to the contractor? If approval is required, then does this mean that once approval is obtained the contractor has no further restrictions? Legal interpretations can vary from State to State, so the agency developing the review requirements must determine what is most appropriate for its State.

10 Establish the QC Requirements

If the QC plan is stipulated by the agency, care should be taken to assure that the plan is not "overkill", i.e., that it is not so complicated or involved that it is viewed as being unworkable. This is not likely to be a concern if the contractor determines the QC plan.

Addressing the QC requirements provides the contractor with the necessary information needed for staffing, procuring laboratory equipment, etc. It is important that the QC plan submitted by the contractor shows some thought and planning on the part of the contractor and does not simply repeat what the agency requires. Some of the QC requirements that the contractor may need to incorporate into the QC plan include:

- Submit the QC plan for review and obtain agency approval.
- Employ qualified technicians.
- Use a qualified laboratory.
- State the properties to be measured and inspected, and the testing frequencies.
- Maintain control charts and state the properties that will be plotted and how often the data will be plotted.
- State the action criteria that will be used to identify "out of control" production by the control charts.
- List the procedures to follow when "out of control" product is identified.
- State who is in charge of correcting "out of control" product.

Specific information on how to establish and use control charts is available from many sources.(8, 9, 10, 11)

10.1. Qualified or Certified Technicians

A comment regarding the use of the terms "qualified" and "certified" technicians is warranted here. FHWA *23 CFR 637B* (6) uses the term "qualified" personnel, as opposed to "certified." One reason that "qualified" was selected is that some States are prohibited by State law to "certify" technicians unless they are State employees. Also, as noted in chapter 3, technician certification usually implies the use of an ongoing recertification program, while technician qualification could be a one–time event. An *AASHTO Draft Standard Recommended Practice for Technician Training and Qualification Programs* (12) has been recently developed. This document indicates that the terms "qualification" and "technician" are meant to be generic descriptions. The *AASHTO QA Guide Specification* (7) uses the term "certified technicians." It is generally understood that technicians must be qualified and that one way to assure this is to require them to have undergone some certification procedure.

If certification is required, most agencies have a certification program. Information on this subject can be found in several documents, three of which are referenced herein. The National Institute for Certification in Engineering Technologies (NICET) sponsors the first two documents in the fields of Highway Materials (13) and Asphalt, Concrete and Soils. (14) The third (15) is an outgrowth of a workshop sponsored by the National Quality Initiative (NQI). This workshop was the result of the requirement contained in FHWA *23 CFR 637B* (6) that *"After June 29, 2000, all sampling and testing data to be used in the acceptance decision or the IA [independent assurance] program shall be executed by qualified sampling and testing personnel."*

Examples

Example QC plans for HMAC and both structural PCC and PCC pavement are provided in appendices B, C, and D. (16)

---

11 Determine Quality Characteristics to Measure for QC

The measurement of some quality characteristics may be more ideally suited for the QC function than for acceptance. For other properties the decision regarding whether to make the quality characteristic part of the QC function, rather than the acceptance function, may be more arbitrary.

In several instances what, in years past, were considered acceptance tests, now are considered better suited for QC tests. One of the reasons for this is that in past specifications the functions of QC and acceptance were not separated. Examples are aggregate gradation from stockpiles for HMAC and slump from fresh mix for PCC, which are now often viewed as QC quality characteristics because they are better early indications as to whether the quality of the product is in control as opposed to being related directly to in–place performance.

Assure that the quality characteristics chosen for testing are suitable for QC purposes. To reiterate, the purpose of QC testing is to measure those characteristics that impact the quality of the product in such a manner that production changes can be made in a timely manner. For example, while it may provide useful information for the agency and contractor, 28–day concrete cylinder strength is not a good QC quality characteristic. By the time this quality characteristic is measured, too much production has occurred to make the strength results useful as a QC tool. If the quality characteristic to be tested is found to be nonresponsive for QC purposes, another characteristic or test method must be found that is appropriate.

11.1. Typical Quality Characteristics for QC Testing

For HMAC, typical quality characteristics that may be tested for QC include:

- Aggregate quality, including fractured faces, sand equivalency, cleanliness, etc.
- Nuclear density.
- Gradation of critical sieve sizes.
- Plant and discharge temperatures.
- Degree of aggregate coating.
- Moisture content of fine aggregate and/or of finished mix.

For PCC, typical quality characteristics that are tested for QC include:

- Aggregate quality.
- Gradation of critical sieve sizes.
- Air content.
- Water–cement ratio.
- Mix temperature.
- Slump.

12

Determinations to Be Made for Each QC Quality Characteristic

Two determinations must be made for each quality characteristic considered for QC testing: (1) the determination of requirements for sampling and testing procedures and (2) the required testing frequency. As noted above, there are several ways that have been successfully used to implement a QC plan. The plan may be stipulated by the agency, chosen by the contractor, or it may be a combination of the two in which the agency chooses some ingredients and leaves others up to the choice of the contractor. The method chosen will impact the way that the QC plan is implemented and the decisions that must be made throughout the implementation of the plan.

13

Evaluate the Available Data

What data are necessary for the agency to consider will be related to what decision has been made regarding which party will develop the QC plan. If the agency decides to stipulate the QC plan requirements, or to stipulate minimum QC requirements, then data will need to be analyzed so that decisions regarding test frequencies and, possibly action or control limits, can be made. If the contractor is to be responsible for developing the QC plan, then it may not be necessary for the agency to evaluate as much data at this point in the process. In an ideal situation in which the contractor is fully responsible for the QC function, the agency would, in theory, not need to analyze any data. However, during the initial implementation of QA specifications, most agencies have not been comfortable giving the contractor total responsibility for developing the QC plan. Therefore, it is likely that the agency will choose to evaluate data for prospective QC quality characteristics testing.

To establish the action limits for control, available data must be analyzed to determine the variability and the ability to hit the "target." Whether these limits are established by the agency or the contractor depends upon the decision regarding whether the agency will stipulate minimum or total QC requirements. If the agency allows the contractor to develop the QC plan, then the plan can be operation–specific, i.e., geared to the contractor's specific plant operation. If the agency decides to stipulate QC requirements, then these requirements will have to be generic in nature so as to apply to many different contractors and different plant operations. The best QC scenario is for the QC plan and the properties tested to be specific to each contractor plant or operation.

If the QC plan is operation–specific, the data should come from prior production of similar product from that operation. In this case, the contractor will be responsible for gathering the data and establishing the operation–specific action or control limits. In general, a measure of the average and the variability must be examined over a period of time and used to establish control/action (and, possibly, warning) limits.

If the QC plan is generic, the agency must evaluate data from a number of typical operations to establish these limits. Published reports of product variability are another source of information that can be used for establishing at least preliminary estimates for action limits. If this approach is taken, the limits will not be as useful since they will not have been developed for each specific plant or operation. For this reason, if the agency does choose to establish initial statewide action or control limits, they should move as quickly as possible toward transferring this responsibility to the contractors so that they can develop operation–specific limits for each plant or operation.

Specification limits should not be used for control/action limits. Control limits are based on the variability of the specific operation or process, and are not derived from the specification limits. If the contractor's process variability does not allow control limits that are completely within the specification limits, then this indicates that the contractor cannot consistently meet the specification requirements.

Since QC is recognized as a contractor or material producer function, the agency should be wary of establishing control or action limits to use on a statewide basis. By necessity, to allow for the many different contractors and plant operations in each State, these limits would have to be set very wide, i.e., conservatively. As such, they may be of little actual benefit in establishing useful control limits for a specific plant operation. However, the agency may choose to initially establish such limits until the industry has become sufficiently knowledgeable for each contractor to take over the full QC function and establish operation–specific limits that are likely to be more restrictive than those initially stipulated by the agency.

13.1. Analysis of Data

The collection and analysis of the QC data should be compatible with the intended use. That is, the data must be collected for the specific quality characteristics in the same manner and under the same general conditions as they will be used. For example, historical data on aggregate gradation from one quarry may not be appropriate for use is establishing control limits for aggregate from a different quarry. Or, historical data for dry aggregate gradations would not be appropriate if the new QC plan called for a washed gradation analysis.

The analysis should involve the statistical properties that will be used to establish action limits. This may be the average, moving average, range, standard deviation, etc. Detailed discussions on the use of each of these measures for QC using control charts are available from many references.(8, 9, 10, 11)

13.2. Use of Historical Data

From where can the necessary data be obtained? Care must be taken when using historical data because they may not always be unbiased. In fact, historical data may frequently be biased.

To be valid, the historical data must have been obtained using a random sampling procedure. That is, the sampling locations should have been selected at random, and not systematically or by the judgment of the inspector. When judgment is used for sample selection, bias can be introduced because there may be a tendency to select a sample location where the material looks questionable in an effort to ensure that "bad" material is not incorporated into the project. On the other hand, there could be a tendency to select a sample location where the material looks particularly good in an effort to ensure that the material is accepted. Either of these will provide a biased estimate of the actual properties associated with past construction.

Another potential problem with historical data is the past process of selecting and testing a second sample when the first sample failed to meet the specification requirements. If the second sample met the specifications, and as a result the first sample was disregarded and its value not recorded, then the historical data will tend to underestimate the actual variability associated with the process being measured. The validity of historical data must be scrutinized thoroughly before deciding to use them to establish QC limits.

14

Determine Whether Sufficient Data Are Available

Are there sufficient data to make the decisions that will form the basis of the QC plan? If the QC plan is operation–specific, the data for the QC properties must be available from the plant operation, and it will be the contractor who should obtain and analyze these data. These data should have been collected in the same manner as that in the QC plan that will be implemented.

This is preferably from a random sampling plan, however, under certain conditions, QC data may be obtained in a selected manner. This is particularly true if a “check” sample is obtained to verify “out of control” or nonconforming product. In this case, all data must be recorded and used in the analysis as to whether the product was truly in or out of control.

If the QC plan is to be generic, the data must come from a number of typical operations and should be randomly sampled. In this case, the data will likely be collected by the agency. These data may come from agency historical records, or may be provided by various contractors. In either case, the same stipulations in the previous paragraph regarding random sampling and collection procedures apply.

If there are not sufficient data, the needed data must be obtained. If sufficient data are available, the next step is to perform the analysis. What constitutes sufficient data may vary from agency to agency. Data should be collected from operations of a number of different contractors, and covering all of the districts or geographic regions of the State. The data should not be just from operations considered to be superior, but should cover a range of operations from the best to those that are considered to be just acceptable. The limits should not be established based on the best operations, but should be a compromise such that they can be achieved by those operations that the agency deems to have been providing acceptable material. It is likely that data from at least 10, and preferably up to 20, projects will be needed for analysis.

15 Obtain the Necessary Data

If sufficient data are not available, there are several potential sources for these data. If the plan is developed by the contractor to be operation–specific, the most appropriate source is from the operation that is to be controlled. However, when the agency decides to initially develop a generic plan, the data should come from several typical operations. An alternative source for these data may be provided from literature reviews. Caution should be observed in this case because there is often a concern as to how applicable data from other places or sources, such as those provided by literature reviews, are to the operation for which the QC plan is being developed. It is important to ascertain that the data from other sources are applicable since the contractor will be held to the limits established from these data.

16 Determine Sampling and Testing Procedures, and Frequency Necessary to Control the Product

The sampling procedures for control, including the point of sampling and the method for selecting the sample, must be established. The quality characteristics, test procedures, and inspection activities that are best related to early indications of process control should be stipulated. These decisions will depend on the type of operation for which the QC plan is being developed. Also, a testing frequency must be established that attempts to create a balance between enough tests to determine if the product is in control but not so many tests as to be impractical. The agency or contractor may decide to use increased testing frequencies initially, and then to reduce these as the operation is shown to be in control.

Normally, the sample should be selected in a random manner. The reasons for this are discussed above and again in chapter 5. The results must be unbiased and all results must be reported. However, there are occasions where nonrandom samples are typically allowed for QC samples. This may happen when an “out of control” result is obtained. The reason for this result could be a sampling or testing problem and therefore not due to a material or production issue. Rather than make an immediate change in the production process, a retest may be taken to verify or dispute the original “out of control” result. When this is done, all test results should be reported and the retest should be noted as such.

16.1. Operation–Specific QC Procedures

Realistically, the QC procedures are different for each operation and, ideally, the decisions should be operation–specific. For example, some operations may require sampling from different locations than others. Likewise, operations with a history of QC problems will require more frequent sampling and testing than operations that typically have had few problems. One of the keys to achieving a balance in testing frequency is to relate the testing frequency to the rate and consistency of production. If the production tends to be continuous and consistent, less frequent testing may be permissible than if there are many interruptions. The contractor should be in the best position to know what tests are the best indicators of control and what frequency is necessary for control.

16.2. Generic QC Procedures

Although the ideal QC procedures are operation–specific, many agencies do not require each contractor to establish these. There are practical reasons that agencies choose to make them generic rather than operation–specific. For instance, an inadequate QC staff to perform the level of testing that should be done may inhibit the contractor. Or the contractor may not want to test more often than the competition because of the impact on staffing and cost. These are some of the reasons that many agencies stipulate at least minimum QC plan procedures. A generic QC plan has potential disadvantages both to the contractor and to the agency. The sampling and testing procedures and the test frequency are stipulated and will not be operation–specific. This is a disadvantage for the contractor that has few occasions of “out of control” product. It also will be a disadvantage to the agency when encountering a contractor that needs more frequent sampling and testing to maintain control.

16.3. Establishing QC Tests and Frequencies

If the agency has opted for contractor–developed, operation–specific procedures, then the choice of testing procedures and frequencies is up to each individual contractor. However, as noted above, many agencies decide to stipulate at least minimum QC testing requirements. The selection of the testing procedures to be required and the testing frequency to stipulate will be done by the joint agency/industry task force that was established during Phase I. While the final decision on the QC requirements to stipulate is made by the agency, input should be sought from the industry representatives on the task force. This input regarding what tests and frequencies are necessary to control an operation should be given careful consideration by the agency when deciding upon QC requirements to stipulate.

Example QC plans, which present suggested QC tests and frequencies for HMAC and both structural PCC and PCC pavement, are provided in appendices B, C, and D. (16)

17 Select Quality Characteristics Most Valid for QC Purposes

As a final check, each quality characteristic that was selected for QC purposes should be reconsidered once data have been collected, analyzed, and reviewed by the agency/industry task force. A final decision is then made whether or not to include the quality characteristic in the stipulated QC requirements. If the decision is to not measure the quality characteristic for QC purposes, then the task force should decide whether the quality characteristic should be considered for possible acceptance testing, or should be eliminated from further consideration.

18 Develop Acceptance Procedures and Requirements

At this point, the QC requirements and quality characteristics for QC sampling and testing have been identified. The QC portion of the specification development process is now ready to proceed to the development of the initial draft QA specification.

CHAPTER 5. ACCEPTANCE PROCEDURES

This chapter continues the discussion of Phase II of the specification development process. This chapter is intended to provide “how to use” best practices in the development or modification of the acceptance plan portion of QA specifications. The steps that are involved in this part of the process are identified in the flowchart in figure 7. The numbers in boxes before the titles of the following sections refer to the corresponding box in the flowchart.

19 Develop Acceptance Procedures and Requirements

There are many important acceptance procedure issues that must be decided upon when developing the acceptance plan and many requirements that can be initiated. As with QC, there is no single prescription that works best in all situations, but there are several that have been effectively used by various agencies.

It is important that the agency determine what it wishes to accomplish with the acceptance plan and its procedures.

- If the primary function is to ensure that the contractors do not totally disregard quality, then the presence of an agency inspector accompanied by a minimal amount of acceptance testing may be sufficient. This limited effort, however, will not really allow the agency to distinguish between good and poor construction and material. To do this will require additional random sampling and testing along the lines of what has traditionally been done or greater.

- If the agency wants a sound statistically based plan that will enable them to determine with a low degree of risk the quality levels that the contractor is providing, then even larger sample sizes will be required.

- If the agency wants to provide sufficient information to use as input into some of the elaborate performance models that are now under development will require considerably more sampling and testing than have traditionally been done by agencies. In this age of competition for limited resources, it seems unlikely that many agencies will be willing to commit to this level of sampling and testing on all of their projects. This level of testing may be limited to selected projects that might be used to help develop agency–specific performance models. However, the calibration and updating of such models would still require an ongoing level of testing that agencies have been unwilling to maintain in the past. It seems unlikely that they will be willing or able to do so in the future.

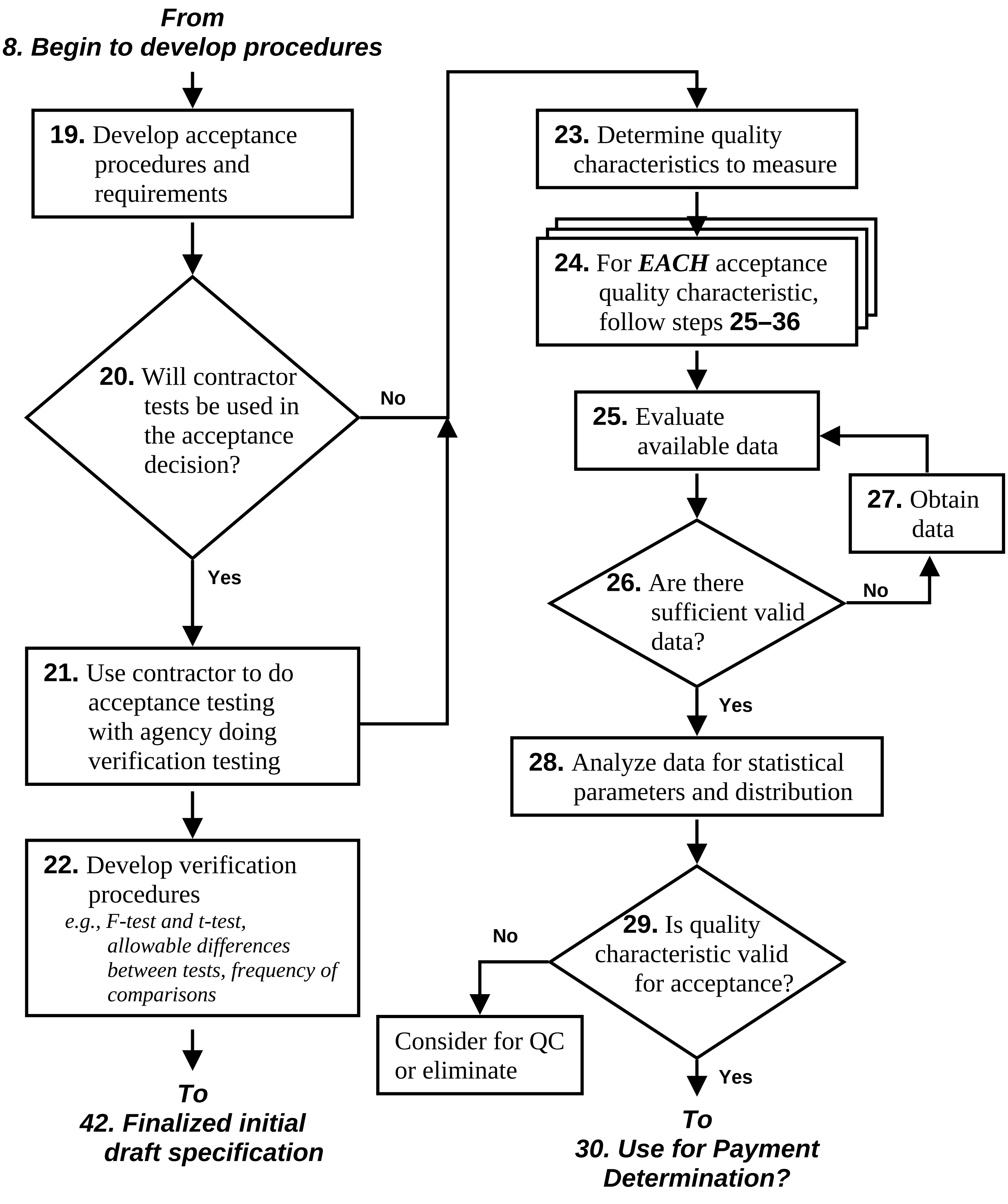


Figure 7. Flowchart for Acceptance Procedures Portion of Phase II

The discussions and examples regarding how to analyze data and develop acceptance plans that are presented in this and the following chapters will help an agency to decide how much sampling and testing it believes is economically justified for its particular situation.

20 Decide Who Will Perform the Acceptance Testing

As part of the acceptance procedures and requirements, one of the first decisions that must be made is "Who is going to perform the acceptance tests?" The answer to this question will influence subsequent decisions and procedures in the acceptance plan. The agency may decide to do the acceptance testing, may assign the testing to the contractor, may have a combination of agency and contractor acceptance testing, or may require a third party to do the testing.

The decision as to who does the testing usually emanates from the agency's personnel assessment, particularly in days of agency downsizing. However, the lack of personnel availability by the agency should not be the sole reason to decide to use contractor acceptance testing, even though this has often been the case. In fact, agencies have sometimes found no significant decrease in agency personnel resulting from the use of contractor acceptance testing. Also, if an agency adopts contractor acceptance testing solely to reduce the agency's staffing needs, then the agency is less likely to follow all the steps, such as developing appropriate validation procedures and conducting pre–implementation training, necessary to successfully implement the QA specification. Furthermore, contractors should never be assigned the responsibility for acceptance testing without being given sufficient preparation time to assume this task, especially in terms of personnel and facilities.

21 Contractor/Third–Party Acceptance Testing

If the agency does the acceptance testing, "business as usual" will be the predominate theme and the next step is to determine what quality characteristics to measure. If the agency does not do the acceptance testing, then it must decide who will perform that function prior to determining what quality characteristics to measure.

Many agencies are requiring the contractor or a third party to do the acceptance testing. As mentioned, this has often come about, at least partially, because of agency staff reductions. What has often evolved is that the contractor is required to perform both QC and acceptance testing. This is one reason that these two functions can become intermingled if care is not taken to assure their separation. If both functions are assigned to the contractor, it is imperative that the difference between the two functions and the purpose for each is thoroughly explained to both contractor and agency personnel. Additionally, if the contractor is assigned the acceptance function, the contractor's acceptance tests must be verified by the agency. Statistically sound verification procedures must be developed that require a separate verification program. There are several forms of verification procedures and some forms are more efficient than others. To avoid conflicts, it is in the best interest of both parties to make the verification process as effective and efficient as practicable.

22 Develop Verification Procedures

If the contractor or a third party acting on behalf of the contractor, such as a consultant, is required to do the acceptance testing, the agency must have a verification procedure to confirm or refute the acceptance test results.

FHWA *23 CFR 637B* (6) states the following:

Quality control sampling and testing results may be used as part of the acceptance decision provided that:

> *(A) The sampling and testing has been performed by qualified laboratories and qualified sampling and testing personnel.*
>
> *(B) The quality of the material has been validated by the verification testing and sampling. The verification shall be performed on samples that are taken independently of the quality control samples.*

The essence of this requirement is a valid and reasonable way to protect taxpayer's interests. An outline of FHWA *23 CFR 637B* requirements is shown in appendix E.

The stated use of QC sampling and testing results for acceptance in FHWA *23 CFR 637B* (6) does not agree with the philosophical approach used in this manual. That is, that QC and acceptance are separate functions and should not be commingled. The reasons for this are presented and discussed in detail in chapters 3 and 4 of this manual. In this manual, contractor tests that are used in the acceptance decision are referred to as acceptance, rather than QC, tests. QC tests are those used by the contractor for the purpose of process control. While it is true that contractors will definitely relate to their processes the results of the acceptance tests, truly beneficial QC tests are those for which results can be obtained during the process so that adjustments can be made to help ensure that the subsequent acceptance tests will meet the requirements of the acceptance plan.

22.1. Definition of Verification

The TRB glossary (2) defines verification as follows:

- ***Verification****—The process of determining or testing the truth or accuracy of test results by examining the data and/or providing objective evidence. [Verification sampling and testing may be part of an independent assurance program (to verify contractor QC testing or agency acceptance) or part of an acceptance program (to verify contractor testing used in the agency's acceptance decision).]*

As noted in chapter 1, some definitions in FHWA *23 CFR 637B* (6) may differ somewhat from those used in this manual. The definitions used here are intended to assure that QC sampling and testing is a separate function from acceptance sampling and testing. However, the need for verification procedures is the same for both sets of definitions. FHWA *23 CFR 637B* (6) uses the term "verification sampling and testing" and defines it as *"Sampling and testing performed to validate the quality of the product."* In this sense, agency verification sampling and testing and agency acceptance sampling and testing have the same underlying function—to validate the quality of the product.

22.2. Independent vs. Split Samples

The TRB glossary (2) contains the following definitions:

- ***Split sample****—A sample that has been divided into two or more portions representing the same material. [Split samples are sometimes taken to verify the acceptability of an operator's test equipment and procedure. This is possible because the variability calculated from differences in split test results is comprised solely of testing variability.]*

- ***Independent sample****—A sample taken without regard to any other sample that may also have been taken to represent the material in question. [An independent sample is sometimes taken to verify an acceptance decision. This is possible because the data sets from independent samples, unlike those from split samples, each contain independent information reflecting all sources of variability, i.e., materials, sampling, and testing.]*

FHWA *23 CFR 637B* (6) requires that *"The verification sampling shall be performed on samples that are taken independently of the quality control samples."* Thus, this procedure does not permit the use of split samples. The need for the use of independent samples as opposed to split samples has been questioned by some agencies.

To understand the difference in the information provided by the two sampling procedures, i.e., split vs. independent samples, an understanding of the concept of components of variability is helpful. Variability can come from many different sources. Statisticians sometimes refer to these variabilities as "errors"—sampling error, testing error, etc. These terms mean sampling variability and testing variability, not mistakes. The sources of variability are combined by the use of the basic measure of variability, called the variance, denoted as σ^2. The sources of variability are combined by adding the variances (not the standard deviations, denoted as σ).

The sources of variability are important when deciding whether to use independent or split samples. The decision depends upon what the agency wants to verify. Independent samples, i.e., those obtained without respect to each other, contain up to four sources of variability: material, process, sampling, and testing method. Split samples contain only testing method variability. These variability components are illustrated in figures 8 and 9.

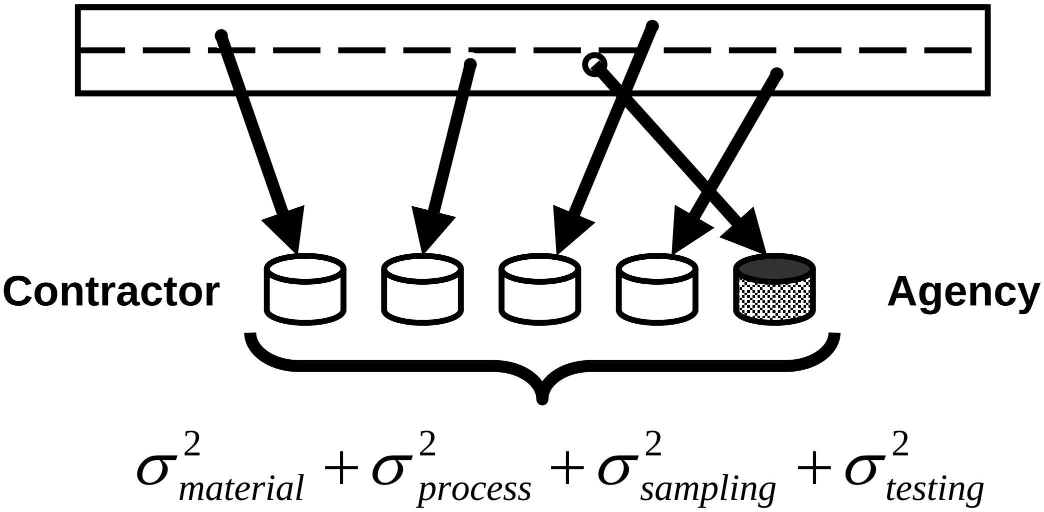


Figure 8. Components of Variance for Independent Samples

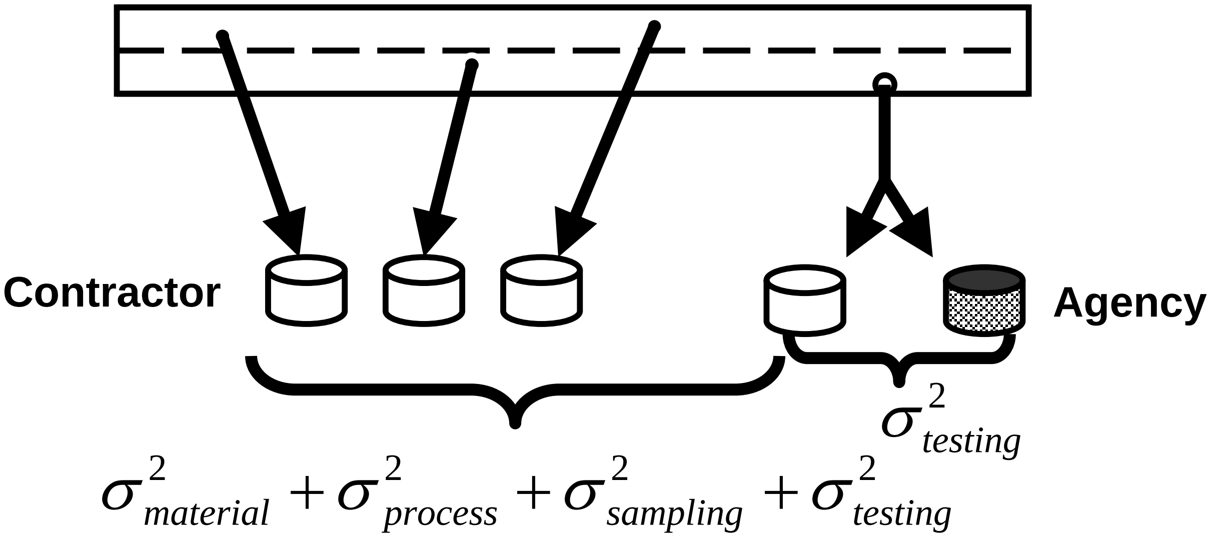


Figure 9. Components of Variance for Split Samples

There has been a considerable amount of confusion between the uses of independent versus split sampling procedures. In an attempt to reduce this confusion, in this manual, the term test method verification refers to the case where split samples are used, while the term process verification refers to the case where independent samples are used.

The statistical implications of these terms extend further than mere definitions. If independent samples are used to verify that two sets of data are statistically similar, then the agency could consider combining the two data sets to make the acceptance decision. Variability issues must be considered when making a decision whether or not to combine the two sets of data. The fact that the data are not shown to be different does not *necessarily* mean that they *are* the same. It simply means that they could not be proven to be different given the sample sizes that were involved. Therefore, it is possible that combining the two sets of data could lead to increased variability. On the other hand, the increased number of values in the combined data set might offset a possible increase in variability. In general, it is probably best to use the agency's verification tests simply for verification, and to use only the contractor's acceptance tests if they compare with the agency's tests.

However, if split samples are used to verify two sets of data, these data should not be combined to make the acceptance decision, even when they were determined to be statistically similar. This is because the two split–sample test results represent essentially the same material, and therefore there is little to no additional information provided by using both results. In fact, using both split–sample test results simply represents a double counting of this particular sample location.

22.3. Verification Sampling and Testing Frequencies

There are no universally accepted verification sampling frequencies. However, like any statistical procedure, the ability of the comparison procedure to identify differences between two sets of results depends on several factors. One of these is the number of tests that are being compared—the greater the number of tests, the greater the ability of the procedure to identify statistically valid differences. A minimum agency rate of 10 percent of the testing rate of the contractor or third party has been used as a rule of thumb.

In practice, the verification testing frequency is usually an economic, rather than statistical, decision. The statistics of the issue will generally call for as many, or more, tests as the agency has the resources to perform. A detailed discussion of the effects of verification testing frequency is presented in the technical report for this project. (17)

22.4. Verification Procedures

22.4.1. Hypothesis Testing and Levels of Significance. Before discussing the various procedures that can be used for test method verification or process verification, two concepts must be understood: hypothesis testing and level of significance. When it is necessary to test whether or not it is reasonable to accept an assumption about a set of data, statistical tests, called hypothesis tests, are conducted. Strictly speaking, a statistical test neither proves nor disproves a hypothesis. What it does is prescribe a formal manner in which evidence is to be examined to make a decision regarding whether or not the hypothesis is correct.

To perform a hypothesis test, it is first necessary to define an assumed set of conditions known as the null hypothesis, H_o. Additionally, an alternative hypothesis, H_a, is, as the name implies, an alternative set of conditions that will be assumed to exist if the null hypothesis is rejected. The statistical procedure consists of assuming that the null hypothesis is true, and then examining the data to see if there is sufficient evidence that it should be rejected. H_o cannot actually be proved, only disproved. If the null hypothesis cannot be disproved (or, to be statistically correct, rejected) it should be stated that we “fail to reject,” rather than “prove” or “accept,” the hypothesis. In practice, some people use “accept” rather than “fail to reject,” although this is not exactly statistically correct.

Example

Consider concrete compressive test cylinders as an example. The null hypothesis might be that the average strength of a concrete bridge deck is 35,000 kilopascals (kPa), while the alternative hypothesis might be that the average strength is less than 35,000 kPa. If three tests are performed—and the test results are 30,300, 31,000, and 31,700 kPa—this would seem to be ample evidence in this simple example that the average strength is not 35,000 kPa, so the null hypothesis would be rejected. The alternative hypothesis, that the average strength is less than 35,000 kPa, would therefore be assumed true.

An important technical point to be aware of is that null hypotheses involve equalities (relationships with “=” signs, e.g., average strength = 35,000 kPa, etc.), while alternative hypotheses involve inequalities (“<”, “>”, or “≠”).

Hypothesis tests are conducted at a selected level of significance, α, where α is the probability of incorrectly rejecting the H_o when it is actually true. The value of α is typically selected as 0.10, 0.05, or 0.01. If for example, $\alpha = 0.01$ is used and the null hypothesis is rejected, then there is only 1 chance in 100 that H_o is true and was rejected in error.

22.4.2. Test Method Verification Procedures. The two procedures used most often for test method verification are the D2S limits and the paired *t*–test.

D2S Limits. This is the simplest procedure that can be used for verification, although it is the least powerful. Because the procedure uses only two test results it cannot detect real differences unless the results are far apart. The value provided by this procedure is contained in many AASHTO and ASTM test procedures. The D2S limit indicates the maximum acceptable difference between two results obtained on test portions of the same material (and thus, applies to only split samples), and is provided for single and multilaboratory situations. It represents the difference between two individual test results that has approximately a 5 percent chance of being exceeded if the tests are actually from the same population.

When this procedure is used for test method verification, a sample is split into two portions and the contractor tests one split–sample portion while the agency tests the other split–sample portion. The difference between the contractor and agency test results is then compared to the D2S limits. If the test difference is less than the D2S limit, the two tests are considered verified.

If the test difference exceeds the D2S limit, then the contractor's test result is not verified, and the source of the difference is investigated.

Example

Suppose that an agency wished to use the D2S limits for test method verification of a contractor's asphalt content determination using the ignition method. *AASHTO T 308–99*, "Determining the Asphalt Binder Content of Hot–Mix Asphalt (HMA) by the Ignition Method," indicates that the D2S limit for two different laboratories is 0.17 percent. So, for a split sample, if the difference between the contractor and agency results is 0.17 percent, or less, the test method would be considered verified. If the difference is greater than 0.17 percent, then the results would be considered different, and an investigation should begin to determine the reason for the difference.

Paired *t*–test. For the case in which it is desirable to compare more than one pair of split–sample test results, the *t*–test for paired measurements can be used. This test uses the differences between pairs of tests and determines whether the average difference is statistically different from 0. Thus, it is the difference *within* pairs, not between pairs, that is being tested. The *t*–statistic for the *t*–test for paired measurements is:

$$t = \frac{\left|\overline{X}_d\right|}{\frac{s_d}{\sqrt{n}}} \qquad (1)$$

where: $\overline{X}_d$ = average of the differences between the split sample test results.

s_d = standard deviation of the differences between the split sample test results.

n = number of split samples.

The calculated *t*–value is then compared to the critical value, t_{crit}, obtained from a table of *t*–values at a level of $\alpha/2$ and with $n - 1$ degrees of freedom. A table of critical *t* values is presented in appendix F. Computer programs, such as Microsoft® Excel, contain statistical test procedures for the paired *t*–test. This makes the implementation process straightforward.

Example

Suppose that an agency wished to use the paired *t*–test for test method verification of a contractor's asphalt content determination using the ignition method. Table 1 shows information on the results of 10 split sample tests that have been conducted.

The *t*–statistic for the differences in table 1 is

$$t = \frac{\left|\overline{X}_d\right|}{\frac{s_d}{\sqrt{n}}} = \frac{0.06}{\frac{0.05}{\sqrt{10}}} = 3.795 \qquad (2)$$

From the table of critical *t*–values in appendix F, for 9 degrees of freedom (i.e., n – 1, or 10 – 1), the critical value for a level of significance of 0.05 (i.e., $\alpha = 0.05$) is 2.262. Since 3.795 > 2.262, the agency would conclude that there is a difference between its results and the contractor's results. The reason for this difference should therefore be investigated.

Table 1. Asphalt Content Data for Paired *t*–test Example

Sample Pair	Contractor	Agency Result	Difference
1	5.65	5.75	+0.10
2	5.45	5.48	+0.03
3	5.50	5.62	+0.12
4	5.60	5.58	–0.02
5	5.53	5.60	+0.07
6	5.51	5.55	+0.04
7	5.78	5.86	+0.08
8	5.40	5.49	+0.09
9	5.68	5.67	–0.01
10	5.70	5.80	+0.10
Average	5.58	5.64	+0.06
Standard Deviation	0.12	0.13	0.05

Recommendation

Test Method Verification Procedure. The comparison of a single split sample by using the D2S limits is simple and can be done for each split sample that is obtained. However, since it is based on comparing only single data values, it is not very powerful for identifying differences when they exist. It is recommended that each individual split sample be compared using the D2S limits, but that the paired *t*–test also be used on the accumulated split–sample results to allow for a comparison with more discerning power. If either of these comparisons indicates a difference, then an investigation to identify the cause of the difference should be initiated.

A more detailed discussion of verification procedures is presented in the technical report for this project. (17)

22.4.3. Process Verification Procedures. Just as there are statistical tests for verification of split sample test results, there are also tests for verification of independently obtained test results. There are two procedures that appear in the *AASHTO Implementation Manual for Quality Assurance*. (16) The tests most often used are the *F*–test and *t*–test, which are usually used together. However, a procedure that compares a single agency test with 5 to 10 contractor tests is also sometimes used. Both of these are discussed below.

***F*–test and *t*–test.** This procedure involves two hypothesis tests, where the H_o for each test is that the contractor's tests and the agency's tests are from the same population. In other words, the null hypotheses are that the variabilities of the two data sets are equal, for the *F*–test, and that the means of the two data sets are equal, for the *t*–test.

When comparing two data sets, it is important to compare both the means and the variances. A different test is used for each of these comparisons. The *F*–test provides a method for comparing the variances (standard deviations squared) of the two sets of data. Differences in means are assessed by the *t*–test. These statistical tests are also commonplace in many computer spreadsheet programs.

The procedures involved with the *F*–test and *t*–test may at first seem complicated and involved. The *F*–test and *t*–test approach also requires more agency test results before a comparison can be made. These reasons may persuade an agency to seek a simpler approach. However**,** the *F*–test and *t*–test approach is the recommended approach because it is much more statistically–sound and has more power to detect actual differences than the second method that relies on a single agency test for the comparison. Any comparison method that is based on a single test result will not be very effective in detecting differences between data sets.

Some of the complexity of the *F*–test and *t*–test comparisons can be eliminated by the use of computer programs. As noted above, many spreadsheet programs have the ability to conduct these tests. In addition, a computer program has been developed specifically for the purpose of making the *F*–test and *t*–test comparisons for process verification testing. (18) This program, DATATEST, conducts both the *F*–test and the appropriate *t*–test for comparing two sets of data. It can conduct the tests at the 0.01, 0.05, or 0.10 levels of significance.

Examples

Appendix F presents a thorough description, along with examples, of both the hand calculations and computer calculations for the *F*–test and *t*–test approach to process verification testing.

Single Agency Test Compared to a Number of Contractor Tests. In this method, a single agency test is compared with 5 to 10 contractor tests. The single agency test result must fall within an interval that is defined from the mean and range of the 5 to 10 contractor test results. The allowable interval within which the agency test must fall is $\overline{X} \pm CR$, where $\overline{X}$ and R are the mean and range, respectively, of the contractor tests, and C is a factor that varies with the number of contractor tests. This is not a particularly efficient approach. This statement, however, can be made for any method that is based on using a single test. Table 2 indicates the allowable interval based on the number of contractor tests. These allowable intervals are based on a level of significance, α, of approximately 0.02.

Table 2. Allowable Intervals for the Single Agency Test Comparison Method

Number of Contractor Tests	Allowable Interval
10	$\overline{X} \pm 0.91R$
9	$\overline{X} \pm 0.97R$
8	$\overline{X} \pm 1.05R$
7	$\overline{X} \pm 1.17R$
6	$\overline{X} \pm 1.33R$
5	$\overline{X} \pm 1.61R$

Examples

More information on this method, including the principles on which it was developed and the magnitude of the difference that is necessary to be identified as significant, are presented in appendix G.

Recommendation

While it is in the *AASHTO Implementation Manual for Quality Assurance,* (16) **THIS METHOD SHOULD NOT BE USED**. This method was developed to be very simple. It suffers from the fact that only a single agency test is used when making the comparison. Any method that relies on a single data value will not be very powerful at detecting differences. This is due to the high variability that is associated with individual, as compared with mean, values.

For example, if the standard deviation for measuring air content in PCC is 0.75 percent, then for a comparison based on five contractor tests, there is only about a 33 percent chance of detecting an actual difference of 2.25 percent between contractor and agency means. The chance only increases to about 57 percent when 10 contractor tests are used. (See appendix G for the development of these values.)

22.5. Power of the Comparison Procedure (i.e., Hypothesis Test)

With any statistical test, the larger the number of test results being compared, the greater the chance of making the correct decision. For the procedures described above, there are operating characteristics (OC) curves available to provide guidance regarding the number of tests needed to achieve a certain probability of detecting a given difference when one actually exists. OC curves plot either the probability of not detecting a difference (i.e., accepting the null hypothesis that the populations are equal) or the probability of detecting a difference (i.e., rejecting the null hypothesis that the populations are equal), versus the actual difference between the two populations being compared. Curves that plot the probability of detecting a difference are sometimes call "power curves" because they plot the power of the statistical test procedure to detect a given difference.

Just as there is a risk of incorrectly rejecting the H_o when it is actually true, the Type I, or α, error, there is also a risk of failing to reject the H_o when it is actually false. This is called the Type II or β error. The "power" is the probability of rejecting the H_o when it is actually false, and is equal to $1 - \beta$. Both α and β are important and are used with the OC curves when determining the appropriate sample size to use.

Figure 10 shows a simple OC curve for the probability of not detecting a difference between two populations. The actual difference between the populations is shown on the horizontal axis, while the probability of NOT detecting the difference is shown on the vertical axis. Three OC curves, for sample sizes of $n = 2$, $n = 4$, and $n =10$, are shown in the figure. For each sample size, when the actual difference between populations is zero (i.e., they are equal) there is a 0.95 (or 95 percent) chance of not detecting a difference. That means that there is a 0.05 (or 5 percent) chance that a difference will be detected when the populations are actually equal. This is the Type I, or α, error.

If we are interested in the ability of the statistical test to identify an actual difference of two units, then figure 10 can be used to identify the Type II, or β, error for this situation. While the Type I, or α, error was the same for each sample size, the Type II, or β, error decreases as the sample size increases. For example, in figure 10, the probability of not detecting a difference of 2 units (depicted on the horizontal axis) is about 0.81 for $n = 2$, about 0.23 for $n = 4$, and essentially 0.0 for $n = 10$.

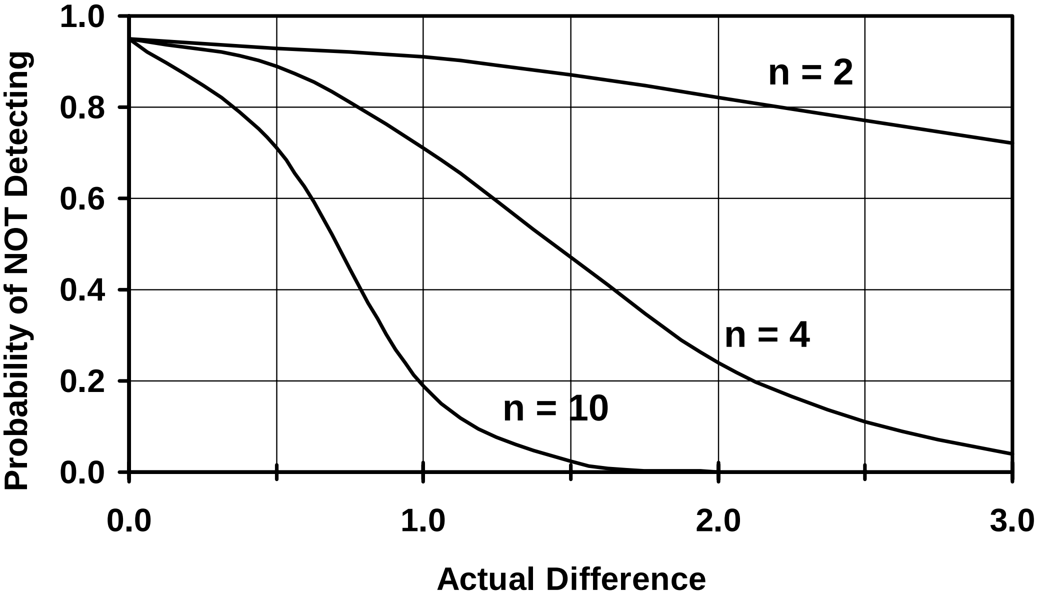


Figure 10. Simple Example of an OC Curve for a Statistical Test Procedure

The frequency of comparison is another decision that must be made. There is no universally accepted frequency. The decision sometimes is related to the outcome of the comparison. For example, the verification may be based on one comparison per lot as long as the contractor's test results are verified. However, the frequency may be increased when the results indicate a statistical difference. When a statistical difference is found, it is important to investigate the difference, find the reason for the difference, and correct the problem if one exists.

Examples

> The OC curves associated with the test method verification methods and the process verification methods discussed above are presented and explained in appendix G. A number of examples are also included in this appendix. An even more detailed discussion of the OC curves is available in the technical report for this project. (17)

---

23 Determine Quality Characteristics to Measure for Acceptance

The measurement of performance–related quality characteristics, i.e., those that relate to in–service performance, is preferred in a QA acceptance plan because they provide an indication that the properties being measured are meaningful. If payment adjustments are made based on the test results for these quality characteristics, these performance–related results can be related to quality through some modeling process. This makes the payment adjustment process rational, and not arbitrary. It is also important to select quality characteristics that can be measured by well–established and reliable test methods. This improves credibility in the selection of the quality characteristic.

Occasions arise in which performance–related quality characteristics either do not exist or require tests that are so sophisticated that they do not have the desirable attribute of providing sufficiently quick results for acceptance testing. For these occasions, surrogate quality characteristics may be chosen in place of a performance–related quality characteristic, but only when absolutely necessary. A surrogate quality characteristic is defined here as one that is measured in place of another quality characteristic or to represent one for which a convenient test does not exist.

An example of the possible use of a surrogate quality characteristic might be for fatigue properties of HMAC in which repeated–load bending beam tests provide a desirable measure of fatigue life. However, bending beam tests are time consuming to run and thus are not considered practical for acceptance purposes. Therefore, the indirect tensile strength might be considered as a surrogate quality characteristic for the quality characteristic fatigue life.

24

Decisions Concerning Each Acceptance Quality Characteristic

Several decisions must be made concerning each acceptance quality characteristic. These decisions include such items as establishing acceptance and/or specification limits, defining acceptable and rejectable quality levels, determining sample size, lot size, sample location, etc. Specific knowledge of each quality characteristic is necessary to make these decisions.

25

Evaluate Available Data

Just as for the QC plan, the sampling and testing procedures for acceptance must be established as well as the specification and/or acceptance limits. To establish these, available data must be analyzed. From where can these data be obtained? Care must be taken when using historical data when developing new acceptance procedures. Historical data may not always be unbiased. In fact, historical data may frequently be biased.

To be valid, the historical data must have been obtained using a random sampling procedure. That is, the sampling locations should have been selected at random, and not systematically or by the judgment of the inspector. When judgment is used for sample selection, bias can be introduced because there may be a tendency to select a sample location where the material looks questionable in an effort to ensure that “bad” material is not incorporated into the project. On the other hand, there could be a tendency to select a sample location where the material looks particularly good in an effort to ensure that the material is accepted. Either of these will provide a biased estimate of the actual properties associated with past construction.

Another potential problem with historical data is the past process of selecting and testing a second sample when the first sample failed to meet the specification requirements. If the second sample met the specifications, and the first sample was therefore disregarded and its value not recorded, then the historical data will tend to underestimate the actual variability associated with the process being measured. The validity of historical data must be scrutinized thoroughly before deciding to use it as the basis for developing new acceptance procedures.

Since the specification and/or acceptance limits will be generic, i.e., agency–wide, the data must not only have been obtained in a manner consistent with their use in the specification, but they must also be broad–based. That means they must have come from production/construction that represents different geographical areas of the State, different contractors with different operations, and projects of different sizes, to mention just some of the considerations. The data must have been obtained by a random sampling procedure and have been sampled and tested in the same manner with the same type equipment as will be required in the new acceptance plan.

26

Determine Whether Sufficient Data Are Available

There are two questions to be answered. First, are the available data valid? For example, have they been obtained in a random, unbiased manner with all results reported? The next question is have the data been obtained from a sufficient number of different contractors and different size projects to provide a description of the quality characteristic of concern? Are there sufficient data to make the necessary decisions, such as estimating statistical parameters and determining an appropriate probability distribution that will form the basis of the acceptance procedure? If the answer is no, the needed data must be obtained. If the answer is yes, the next step is to analyze the data.

There is not a single answer to what constitutes a sufficient amount of data. What constitutes sufficient data may vary from agency to agency. Data should be from the operations of a number of different contractors, and should cover all of the districts or geographic regions of the State. The data should not be just from operations considered to be superior, but should cover a range of operations from the best to those that are considered to be just acceptable. The specification limits that will be established from these data should not be based on just the best operations, but should be a compromise such that they can be achieved by those operations that the agency deems to have been providing acceptable material. It is likely that data from at least 10, and preferably up to 20, projects will be needed for analysis.

27 Obtain Data if Necessary

What happens when the data are not sufficient to make the decision on statistical parameters, etc.? From where can these additional data be obtained?

There are several potential sources for the data. The literature review that was undertaken earlier is one potential source for additional data. However, there is often a concern as to how applicable data from other places are to the location for which the acceptance plan is being developed. It is important to ascertain that the data from other sources are applicable since contract payment may be determined from the outcome of the data analysis. Another data source offers a solution to the problem of applicability. This involves gathering new data from ongoing projects on an agency–wide basis. The advantage of this approach is that the data may be viewed as more appropriate than those from a literature review. The disadvantage is that it will be more time consuming to collect data in this manner. If new data must be collected to perform the evaluation, the specification development schedule should then be reexamined to determine whether or not it should be modified and/or extended.

28 Analyze Data for Statistical Parameters and Distribution

Once there are sufficient data, they must be analyzed to determine the appropriate parameters and distributions to use when developing the acceptance procedures. The analyses should determine the appropriate probability distributions to use to represent the material or process. The analysis should also develop estimates for population parameters—mean, standard deviation, variance, and possibly bias, skewness, and kurtosis—on a project–by–project basis. "Typical" values, particularly for process standard deviation, should also be developed. All of this information may be useful in establishing specification limits and defining quality levels. The use of computerized statistical programs, which typically include histograms and normal probability plots as well as calculation of statistical measures, is recommended.

28.1. Determine Appropriate Probability Distribution for the Data

Attribute acceptance plans do not require that the population from which the data are obtained be from a normal distribution, whereas the use of a variables acceptance plan usually requires that the data be from an approximately normal distribution. Ascertaining that this assumption is correct is important. Visual observation of data histograms, plotting data on normal probability paper, examining skewness and kurtosis values, and statistical goodness–of–fit (GOF) tests, such as the Chi Square and Kolmogorov–Smirnov tests, are some methods available to check the normality assumption.

Although most construction materials cases have been shown to be approximately normal, there are procedures available to assess skewed distributions when they are known to occur. Two ways of addressing skewness are:

- To use an attribute acceptance plan, which is not as efficient in the relationship between sample size and risk as a variables plan.
- To use averages of test results rather than individual values. Averages of data have been proven to be approximately normally distributed even when taken from a non–normal population.

Examples

The reader is referred to appendix H for a description and examples of some simple methods for assessing the normality of a set of data. More statistically rigorous goodness–of–fit tests are not discussed herein, but their procedures are available in numerous statistical texts.

28.2. Determine Appropriate Process Variability

The first question that must be answered is "What variability will be used for the typical variability on which to base the acceptance requirements?" There are several issues that must be addressed when answering this question, and they are discussed in the following sections.

28.2.1. Which "Project" Variability Is Appropriate? The first, and perhaps most important, issue is to develop a value for project variability that is consistent with the way in which a lot will be defined under the new QA acceptance plan.

Combined "Project" Standard Deviation. It is generally appropriate to combine all test results from a given past project, and to calculate an overall standard deviation value for the combined data, ONLY if in the future the entire project will be used as a single lot for payment determination. Such a decision to use the entire project as the lot, however, assumes that the results from all of the various paving days on the project can be combined to form a single normal population. This may not always be a good assumption in light of the fact that weather changes and process adjustments are frequent occurrences over the life of a typical paving project. In the past, some agencies have calculated combined overall project standard deviation values and then used these to establish specification limits where there is lot–by–lot, rather than total project, acceptance. This is NOT correct, and should not be done.

Typical "Within–Lot" Standard Deviation for a Project. If the new 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. In such a case, it is NOT appropriate to combine all test results from a project and then to calculate a standard deviation for these combined data. The individual standard deviation values for each lot must be calculated and then these lot standard deviations are pooled to get a typical "within–lot" standard deviation for the process.

How the individual lot standard deviation values are averaged or combined depends upon the number of test results there are for each lot. From a statistical standpoint, the number of test values for each lot is known as the sample size, or *n*. In this case, sample size refers not to the quantity of material that comprises the individual test portions, but to the number of test results on which the lot information is based. That is, if there are four tests from each lot, then the sample size is $n = 4$.

Sample Standard Deviation. In statistical terms, the sample standard deviation is referred to as *s*. The individual lot (or *sample* in statistical terms) standard deviations cannot directly be used to estimate the within–lot process (or population in statistical terms) standard deviation. This is true because the sample standard deviation is a biased estimator for the population standard deviation. Therefore, if the individual sample standard deviations are used to estimate a population standard deviation, then a correction factor MUST be applied to adjust for the bias in the estimate. This correction factor is applied to the arithmetic average (or mean in statistical terms) of a number of individual lot standard deviation values, and the sample size should be the same for each of the individual lots. Because of these limitations, it is rarely appropriate to estimate the typical within–lot project standard deviation based on the individual lot standard deviations.

Warning!

It is NOT correct to average individual lot standard deviation values to get a typical project standard deviation. The sample standard deviation is a biased estimator of the population standard deviation. It is acceptable to use the lot standard deviations to estimate the population standard deviation ONLY if the sample size is the same for each lot and if the appropriate correction factor is applied. **The following method that uses the sample variances is the correct and recommended method.**

Sample Variances. The preferred method for estimating the within–lot process standard deviation is by using the lot variances rather than standard deviations. This is true because the sample variance is an unbiased estimator of the population variance. Therefore, the individual lot variances can be "pooled" to provide an unbiased estimate for the within–lot process variance. The square root of the pooled project variance will then be an unbiased estimate for the within–lot process standard deviation. The pooled variance is a weighted average based on the sample sizes associated with the individual lot variances. In statistical terms, if we assume that our individual lot values are from the same population or from different populations having equal variances, then the individual lot variances can be pooled to give an estimate for the within–lot process variance. The formula for this estimate, if there are *k* individual lots, is

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{n_1 + n_2 + \cdots n_k - k} \tag{3}$$

where: s_p^2 = pooled estimate for the within–lot process variance.

s_i^2 = variance for lot i, where $i = 1, 2, \ldots k$.

n_i = number of values for lot i.

k = number of lots in the project.

The pooled standard deviation is then simply the square root of the pooled variance. Although this equation assumes sampling from populations having equal variances, it is generally believed to be adequate for this application when developing acceptance plans.

Example

The test results from a past project are shown in table 3.

Table 3 indicates that when all 40 of the individual test results are combined into one data set, the variability, as indicated by the standard deviation, is 0.69. However, when the standard deviations of the 10 individual lots are pooled using equation 5–3, the "within–lot" process standard deviation is 0.61.

If an agency decides that it will base acceptance on a project–sized lot, then the correct standard deviation to represent this project is 0.69, which is based on a sample size of n = 40 tests in the project. However, if the agency decides that acceptance will be on an individual lot basis, then the approximate standard deviation to represent a typical lot on this project is 0.61, with a sample size per lot of n = 4.

Table 3. Example Summary of Test Results for a Project

Lot	Test Results	*n*	Lot		
			Mean	Std Dev	Variance
1	4.6, 5.3, 5.5, 4.8	4	5.05	0.420	0.176
2	6.0, 5.7, 5.1	3	5.60	0.458	0.210
3	5.2, 3.7, 4.2, 5.0	4	4.53	0.699	0.487
4	6.3, 6.1, 4.9, 6.0, 5.3	5	5.72	0.593	0.352
5	5.2, 5.0, 3.6	3	4.60	0.872	0.760
6	5.8, 4.9, 4.5, 5.5	4	5.18	0.585	0.342
7	4.9, 4.7, 3.5, 4.6	4	4.43	0.629	0.396
8	5.9, 5.6, 4.2, 5.5, 4.7	5	5.18	0.705	0.497
9	5.9, 5.7, 4.4, 5.6	4	5.40	0.678	0.460
10	4.4, 4.6, 5.0, 4.8	4	4.70	0.258	0.067
	Individual Tests	40	5.1	0.69	
	Pooled for 10 Lots	10		0.61	0.371

28.2.2. Selecting a "Typical" Process Variability. After determining typical within–lot process variabilities for various projects, it can then be determined if the data from contractors are reasonably consistent, or whether some have appreciably lower or higher variabilities than others, whether some meet the specification requirements more often than others, etc. This is important when trying to select a "typical" within–lot variability for the overall process (or, process variability). For instance, the typical process variability should not be set for the most or least consistent contractor.

As noted above, the data must come from a number of different projects as well as different contractors. The number of projects to consider will vary depending upon the number of contractors that work in the State, the number of different geological regions in the State, and how much the process variabilities differ among projects and contractors.

Once the project variability data are available, a decision must be made regarding what variability to use as the "typical" process variability. This typical variability will then be used to establish specification limits. There is no single "correct" way to decide upon the typical variability to use. An example may help to clarify some of the factors involved in the decision.

Example

Suppose that a highway agency has collected data from 10 past projects that it considered acceptable, and determined the results shown in table 4.

The agency could decide to select 1.65 as the "typical" process standard deviation value (measure of process variability) since this value is "capable" of being achieved. On the other hand, the agency could select 3.20 since this value was obtained on a project that the agency had apparently considered acceptable. It is probably not appropriate to select either the best (smallest) variability or the worst (largest) variability as the "typical" variability. An agency cannot reduce variability by simply specifying it, particularly if it has been shown that contractors, in general, have not been able to consistently meet that variability value. It is probably also not a good practice to base acceptance plan decisions on the worst contractor results.

Therefore, the agency would probably wish to select the typical process variability value based on consideration of all of the past project data rather than just a single best or worst project. The agency might order the standard deviation values from smallest to largest. This yields: 1.65, 2.03, 2.05, 2.12, 2.20, 2.35, 2.51, 2.71, 2.84, 3.20. A subjective decision could then yield several possible "typical" values. For instance, 2.51 might be selected since 7 of 10 projects had this value or less, and because there was a fairly large gap to the next higher value (2.71). Similarly, 2.84 might be selected because of the very large gap between this value and the largest value of 3.20. Other subjective choices are possible.

There is no single "correct" way to decide on the typical value for process variability. The agency should consider various options and select the method with which it is most comfortable for the given project data.

Table 4. Example Project Variability Results

Project	**Project Standard Deviation**	**Project**	**Project Standard Deviation**
A	1.65	F	2.51
B	2.05	G	3.20
C	2.84	H	2.20
D	2.12	I	2.71
E	2.35	J	2.03

28.3. Consider Target Value Miss

The typical standard deviation value that is selected serves as a measure of variability within the process, i.e., the “within–process” variability, for a typical contractor on a typical project. This standard deviation will be used to help decide upon specification limits for the acceptance plan. Another factor that needs to be considered in addition to this within–process variability is the capability of contractors to center their processes on the target value. This may be an even more difficult task than deciding on a typical within–process standard deviation.

28.3.1. Combined Typical Standard Deviation. Many possible quality characteristics have target values about which two–sided specification limits will be established. As shown later in this manual, the identified typical process standard deviation can be used to establish these specification limits. The agency, however, must decide whether or not a typical contractor can be expected to always be able to center its process exactly on the target value. If the agency believes this to be possible, then the typical process standard deviation that was developed from the analyses in previous sections is the correct one to use when setting the specification limits. If, on the other hand, the agency believes that a typical contractor’s process mean may vary somewhat about the target value, then it will be necessary to consider this fact when developing specification limits.

It is not reasonable to assume that a contractor can always set its process mean precisely on the target value. Differences in materials, weather conditions, conditions of the existing pavement, and other factors may lead to a contractor’s process mean occasionally missing the target value in spite the contractor’s best efforts to hit it. If current technology does not allow for the process mean to always be “on target,” then the agency needs to consider this when establishing specification limits. Since this target miss will add additional variability to the within–process variability, this will lead to wider specification limits than those established based strictly on the typical process standard deviation.

What is being discussed here is not the case where a contractor, for whatever reason, chooses to intentionally center its process at some point other than the target value. If a contractor chooses to do this, then the contractor must bear any potential acceptance risks associated with its decision. On the other hand, failure to consider that current technology may not be adequate to allow the contractor to always hit the target with all of its processes places an unfair risk on the contractor.

The proper way to address the issue of “target miss,” is to determine how variable the actual process means are about the target value. This variability regarding where the process will be centered, call it “process center variability,” can then be combined with the previously determined typical within–process variability to obtain the correct standard deviation value for use in establishing specification limits.

The “process center variability” and the “within–process variability” can be combined simply by adding their associated variances, NOT their standard deviations. This assumes that the amount of process variability is independent of where the process is centered, an assumption that seems reasonable, particularly as long as the target miss is not very large. Note that it is NOT correct to add the two standard deviations. The two variances must be added to get a combined variance.

The square root of this combined variance is then the correct combined standard deviation value to use. This relationship is shown in the following equations.

$$\hat{\sigma}^2_{combined} = \hat{\sigma}^2_{center} + \hat{\sigma}^2_{process} \tag{4}$$

$$\hat{\sigma}_{combined} = \sqrt{\hat{\sigma}^2_{combined}} \tag{5}$$

where: $\hat{\sigma}^2_{center}$ = estimated process center variance.

$\hat{\sigma}^2_{process}$ = estimated within–process variance.

$\hat{\sigma}^2_{combined}$ = estimated combined process center and within–process variance.

$\hat{\sigma}_{combined}$ = estimated combined standard deviation.

The true answer as to how much process center variability exists is extremely difficult to obtain. The decision may require some engineering judgment since an analysis of data is not likely to be able to yield a clear answer to the question.

One reason that it is difficult to answer this “target miss” question from project data is that the agency never knows with certainty where the contractor intended to center its process. A contractor with particularly low variability could, for a number of reasons, choose to center its process at a point other than the target value and still plan to meet the specification requirements based on its low variability. It will also not be possible to determine from project data whether or not the contractor’s process mean was constant throughout the project or whether for any of a number of reasons it was changed during the course of the project. Any “target miss” analysis will therefore require some assumptions on the part of the agency. Each individual agency must decide, based on its experience, what assumptions it believes are appropriate.

28.3.2. Assuming a Constant Process Throughout the Project. If the agency developed its typical standard deviation by combining all test results from the project, i.e., it decided to use the total project for the acceptance lot, then the agency has already assumed that the contractor’s process remains constant throughout the project. If this assumption is made, then it is possible to use data from a number of projects to estimate the “target miss” variability. As mentioned in the previous discussion for establishing a typical process variability value, it seems unlikely that a contractor’s process will be identical throughout the life a large project. This is the reason that an agency would choose to use the within–lot variation to establish the typical process standard deviation.

However, if the agency does assume a constant process throughout a project, then the mean value for all of the lot means on the project will be a good estimate of where the process was centered for the project. The agency could then obtain a large number of project “target misses” and analyze these to determine the variability associated with missing the target value. An example will help to illustrate how this could be done.

Example

Suppose an agency is willing to assume that a contractor's process remains constant throughout an entire project. The agency could then determine an "average target miss" value for a specific project by subtracting the target value from the average of all of the test results on the project.

Assume that the agency has obtained the following "average target miss" values from 13 past projects:

–0.30, –1.28, +0.24, +1.28, +1.20, +1.73, –2.18, –0.23, +1.10, –1.09, –0.69, –1.69, +1.85.

The mean for these 13 project "target misses" is 0.00. This indicates that, on the average, contractors are probably aiming for the target. The standard deviation for these 13 "target misses" is 1.348, with a corresponding variance of 1.817. These values represent the variability associated with hitting the target value.

Now assume that the agency had previously selected the typical process variability to be represented by a standard deviation of 2.20, with a corresponding variance of 4.84. The combined standard deviation to use for establishing specification limits can then be calculated using equations 6 and 7:

$$\hat{\sigma}^2_{combined} = 1.817 + 4.840 = 6.657 \quad (6)$$

$$\hat{\sigma} = \sqrt{6.657} = 2.58 \quad (7)$$

Caution

The method described in the above example applies ONLY if the agency has decided to assume that the contractor's process remains constant throughout the entire duration of a project. If it is believed that the contractor's process mean may vary from lot–to–lot within a project, which seems likely, then the above approach is NOT appropriate.

28.3.3. Assuming Process Not Constant Throughout Project. If the agency does not believe that the contractor's process is constant throughout the life of a project, as would typically be the case when the agency has decided to use lot–by–lot acceptance, then the procedures in the previous example would NOT be appropriate. In this case there is no easy way to determine a typical "target miss" variability since there is no way to know how much of the lot–to–lot variation in sample means is from the natural variation of the sampling process and how much is due to misses, changes, or adjustments in the contractor's target mean during the project. To address this situation, the agency must make some assumptions. Which assumptions are made will depend upon the individual agency and what it believes to be the most reasonable for its State, contractors, and processes.

One possibility might be to calculate a standard deviation based on combining all of the project data into one data set. While this was not recommended above in the discussion on how to establish a process standard deviation to use with lot–by–lot acceptance, this approach will provide a larger standard deviation value that includes the lot–to–lot variation in lot means. A decision to use this approach assumes that any “target miss” variation within the project will be accounted for when all the test results are combined. The various project standard deviations could then be pooled, using their corresponding variances, to arrive at a typical process standard deviation that attempted to include the possible “target miss” variability.

Another possible approach would be for the agency to use some experience, engineering judgment, and knowledge of the process to develop a reasonable estimate for the “target miss” variability. The agency could base its decision on its past experience as well as discussions with contractors in the State. A very simple example shows how an agency might arrive at a value for “target miss” variability.

Example

Suppose that an agency wishes to determine a “target miss” standard deviation for a particular quality characteristic. Based on the judgment of experienced agency personnel and discussions with contractors, the agency believes that most of the time a contractor can control its process mean to within ±1.5 units of the target mean.

Next, the agency might decide to assume that “most” of the time can be represented by ±2 “target miss” standard deviations. This would correspond to about 95 percent of the time. Under this assumption, the “target miss” standard deviation would be about 1.5 ÷ 2 = 0.75 units. If the agency had previously decided upon a “within–process” standard deviation of 2.75 units, the combined standard deviation for developing specification limits can be calculated using equations 8 and 9.

$$\hat{\sigma}^2_{combined} = 0.75^2 + 2.75^2 = 8.125 \tag{8}$$

$$\hat{\sigma} = \sqrt{8.125} = 2.85 \tag{9}$$

28.4. Identification and Explanation of Data Anomalies

Anomalies in the data should be identified and, if possible, explained. There may be something in the contractor’s operation, the sampling and testing, etc., that leads to the anomaly, and should, therefore, be considered for inclusion or exclusion in the acceptance plan. For example, if data from one project are quite different from the other projects it may be determined that there were special circumstances, such as nighttime paving in cold weather, involved on this project.

The agency would then need to decide whether or not it wished to consider this project when establishing the new acceptance plan requirements.

28.5. An Ongoing Process

As seen from the previous sections, there is no single, clear–cut “correct” method to establish the typical standard deviation value to use when setting specification limits. Each of the possible methods requires some assumption or assumptions by the agency. Each individual agency must decide which of these assumptions it believes are most appropriate for its given situation. Whichever method the agency uses to determine the standard deviation that it uses to set specification limits, the agency should not consider the process to be finished. Once the new acceptance plan is implemented, the agency must continue to collect and to monitor data from projects to verify that the assumptions that were made when developing the acceptance plan were appropriate. It is important that the monitoring data be obtained in the same fashion as the original data that were used to establish the initial typical standard deviation value for the process. The agency must then be willing to modify its typical standard deviation value and corresponding specification limits if the additional project data indicate that it is necessary.

29

Select Best Quality Characteristics for Acceptance Testing

As a final check, each quality characteristic that was selected for acceptance purposes should be reconsidered once data have been collected, analyzed, and reviewed by the agency/industry task force. A final decision is then made whether or not to include the tests to measure each quality characteristic in the stipulated acceptance requirements. If the decision is to not measure the quality characteristic for acceptance purposes, then the task force should decide whether the quality characteristic should be considered for possible QC testing, or should be eliminated from further consideration.

CHAPTER 6. ACCEPTANCE AND PAYMENT PROVISIONS

This chapter continues the discussion of Phase II of the specification development process. This chapter is intended to provide “how to use” best practices in the development of procedures for determining the acceptability of and payment for materials and construction. The steps that are involved in this part of the process are identified in the flowchart in figure 11. The numbers in boxes before the titles of the following sections refer to the corresponding box in the flowchart.

30 Decide if the Quality Characteristic Will Be Used for Payment Determination

The last chapter presented the steps in determining what quality characteristics should be measured as part of the acceptance decision. It must next be decided if each of the characteristics to be measured will be used in the payment factor determination. If the answer is yes, the next step is to determine the quality measure to be used. If not, the property may be used as a screening test on a pass/fail basis.

31 Use of a Screening (Pass/Fail) Test for Acceptance Quality Characteristics

A screening test is one for which the results can be determined quickly enough to prevent “non–conforming” material from being incorporated into the project. A pass/fail criterion is often used with this procedure. The advantage of a screening test is that it can prevent poor quality material from being incorporated into the project, and thus does not require a payment relationship. A potential disadvantage of this type of requirement is that, if a small sample size is used, there is an increased risk of making an incorrect decision. This is illustrated in an example later in this chapter.

In practice, either the agency or the contractor may run screening tests. The agency would use the tests as an acceptance function, while the contractor may choose to use screening tests as a QC function. Either way, the intent is to keep nonconforming material from being incorporated into the project. With the decrease in on–site agency personnel, more screening tests are being performed by the contractor. If this is done, there may be a requirement that agency personnel either “witness” the testing or examine the results on the test reports. Since these tests are pass/fail tests, typically no payment determination is made for nonconforming product. Instead, if this occurs, the product is not incorporated into the project. An exception to this general rule is discussed in case study 1 in chapter 9.

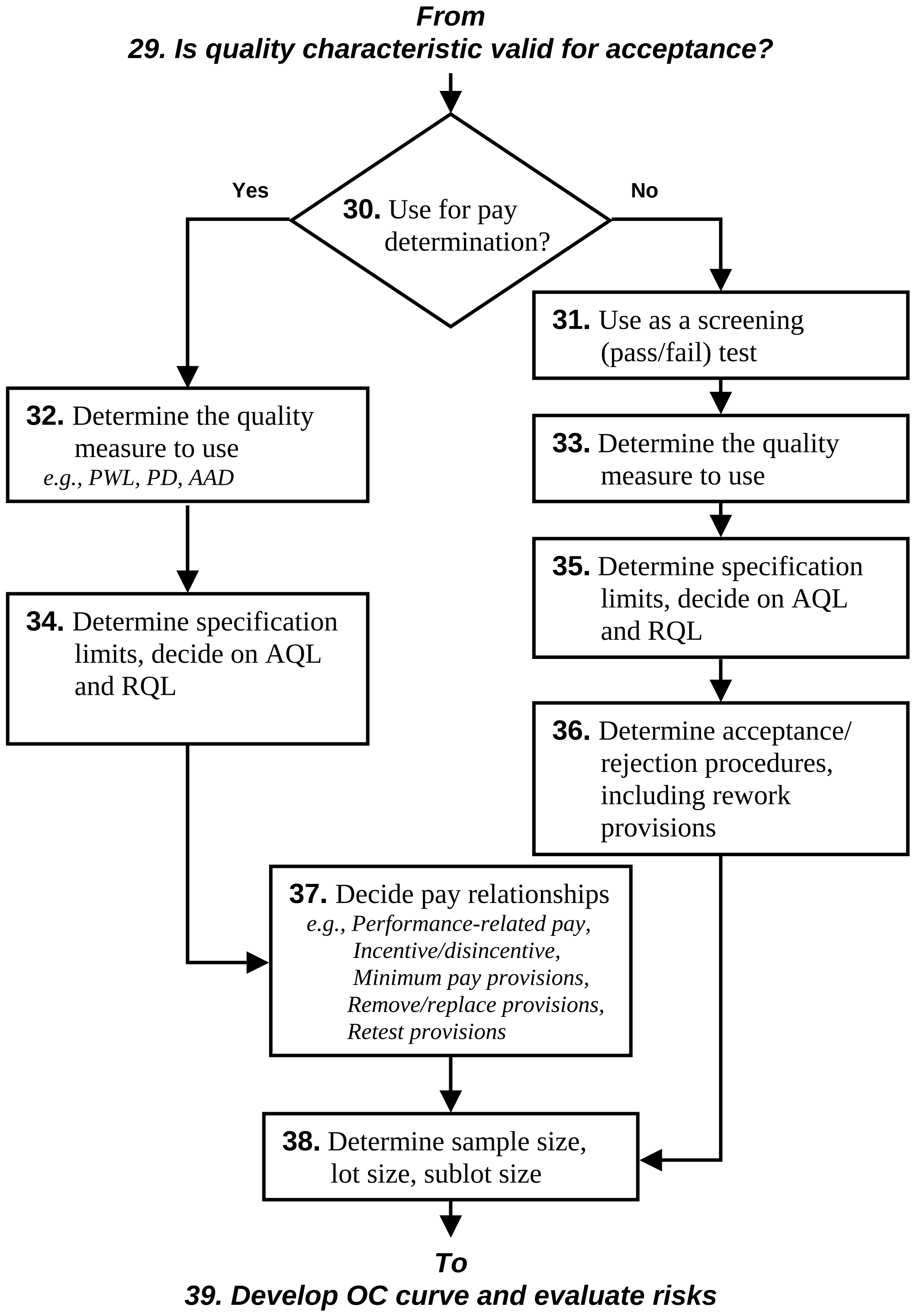


Figure 11. Flowchart for Acceptance and Payment Portion of Phase II

32

Payment Quality Characteristics: Determine the Quality Measure to Use

If the quality characteristic is to be used for payment determination, the quality measure to be related to the payment must be decided upon. There are several quality measures that can be used. In past acceptance plans, the average, or the average deviation from a target value, was often used as the quality measure. However, the use of the average alone provides no measure of variability, and it is now recognized that variability is often an important predictor of performance.

Several quality measures, including percent defective (PD) and percent within limits (PWL), have been preferred in recent years because they simultaneously measure both the average level and the variability in a statistically efficient way. Other measures that have been used by some agencies include the average absolute deviation (AAD) and the moving average. An additional measure that may be considered by some agencies is the conformal index (CI). Some of these measures are more discriminating than others, and the choice of the most effective measure can translate directly into economic savings, due either to a reduced inspection or testing effort or to a lesser amount of poor product accepted, or to both.

32.1. The PWL Quality Measure

The TRB glossary (2) includes the following definition (where LSL and USL represent lower and upper specification limits, respectively):

- ***PWL****—also called* **percent conforming**. *The percentage of the lot falling above the LSL, beneath the USL, or between the USL and LSL.[PWL may refer to either the population value or the sample estimate of the population value. PWL = 100 – PD.]*

This quality measure uses the sample mean and the sample standard deviation to estimate the percentage of the population (lot) that is within the specification limits. This is called the PWL method, and is similar in concept to determining the area under the normal curve.

In theory, the use of the PWL (or PD) method assumes that the population being sampled is normally distributed. In practice, it has been found that statistical estimates of quality are reasonably accurate provided the sampled population is at least approximately normal, i.e., reasonably bell shaped and not bimodal or highly skewed. Normality tests and their application are discussed in appendix H.

32.1.1. Estimating PWL. Conceptually, the PWL procedure is based on the normal distribution. The area under the normal curve can be calculated to determine the percentage of the population that is within certain limits. Similarly, the percentage of the lot that is within the specification limits can be estimated. Instead of using the *Z*–value and the standard normal curve, a similar statistic, the quality index, *Q*, is used to estimate PWL. The *Q*–value is used with a PWL table to determine the estimated PWL for the lot.

A sample PWL table is shown in table 5. A different format for a table relating *Q* values with the appropriate PWL estimate is shown for a sample size of $n = 5$ in table 6. A more complete set of PWL tables in this format, for sample sizes from $n = 3$ to $n = 30$, is available.(18) Another way of relating *Q* and PWL values is presented in table 7. In this table a range of *Q* values is associated with each PWL value. This table was developed by an agency such that any estimated PWL is rounded up to the next integer PWL value. Other possible rounding rules could be used to develop similar tables. The rounding rule in table 7 is the one that is most favorable to the contractor since it rounds any PWL number up to the next whole number. For example, 89.01 is rounded up to 90.00 in table 7.

32.1.2. Calculation and Rounding Procedures. As the previous paragraph illustrates, the calculation procedures and rounding rules can influence the estimated PWL value that is obtained. This can become a point of contention, particularly if the payment determination is based on the estimated PWL value. It is therefore important that the agency stipulate the specific calculation process, including number of decimal places to be carried in the calculations, as well as the exact manner in which the PWL table is to be used.

For example, in table 5, is the PWL value to be selected by rounding up, rounding down, or by linear interpolation. Each of these will result in a different estimated PWL value. For instance, if the sample size is $n = 5$, and the calculated Q value is 1.18, the estimated PWL values for rounding up, rounding down, and interpolating would be 89, 88, and 88.5, respectively.

32.1.3. The Quality Index and PWL. The *Z*–statistic that is used with the standard normal table, an example of which is shown in table 8, uses the population mean as the point of reference from which the area under the curve is measured:

$$Z = \frac{X - \mu}{\sigma} \tag{10}$$

where: Z = the *Z*–statistic to be used with a standard normal table (such as table 8).
X = the point within which the area under the curve is desired.
μ = the population mean.
σ = the population standard deviation.

The statistic *Z*, therefore, measures distance above or below the mean, μ, using the number of standard deviation units, σ, as the measurement scale. This is illustrated in figure 12.

Table 5. Quality Index Values for Estimating PWL

PWL	***n* = 3**	***n* = 4**	***n* = 5**	***n* = 6**	***n* = 7**	***n* = 8**	***n* = 9**	***n* = 10 to 11**
100	1.16	1.50	1.79	2.03	2.23	2.39	2.53	2.65
99	–	1.47	1.67	1.80	1.89	1.95	2.00	2.04
98	1.15	1.44	1.60	1.70	1.76	1.81	1.84	1.86
97	–	1.41	1.54	1.62	1.67	1.70	1.72	1.74
96	1.14	1.38	1.49	1.55	1.59	1.61	1.63	1.65
95	–	1.35	1.44	1.49	1.52	1.54	1.55	1.56
94	1.13	1.32	1.39	1.43	1.46	1.47	1.48	1.49
93	–	1.29	1.35	1.38	1.40	1.41	1.42	1.43
92	1.12	1.26	1.31	1.33	1.35	1.36	1.36	1.37
91	1.11	1.23	1.27	1.29	1.30	1.30	1.31	1.31
90	1.10	1.20	1.23	1.24	1.25	1.25	1.26	1.26
89	1.09	1.17	1.19	1.20	1.20	1.21	1.21	1.21
88	1.07	1.14	1.15	1.16	1.16	1.16	1.16	1.17
87	1.06	1.11	1.12	1.12	1.12	1.12	1.12	1.12
86	1.04	1.08	1.08	1.08	1.08	1.08	1.08	1.08
85	1.03	1.05	1.05	1.04	1.04	1.04	1.04	1.04
84	1.01	1.02	1.01	1.01	1.00	1.00	1.00	1.00
83	1.00	0.99	0.98	0.97	0.97	0.96	0.96	0.96
82	0.97	0.96	0.95	0.94	0.93	0.93	0.93	0.92
81	0.96	0.93	0.91	0.90	0. 90	0.89	0.89	0.89
80	0.93	0.90	0.88	0.87	0.86	0.86	0.86	0.85
79	0.91	0.87	0.85	0.84	0.83	0.82	0.82	0.82
78	0.89	0.84	0.82	0.80	0.80	0.79	0.79	0.79
77	0.87	0.81	0.78	0.77	0.76	0.76	0.76	0.75
76	0.84	0.78	0.75	0.74	0.73	0.73	0.72	0.72
75	0.82	0.75	0.72	0.71	0.70	0.70	0.69	0.69
74	0.79	0.72	0.69	0.68	0.67	0.66	0.66	0.66
73	0.76	0.69	0.66	0.65	0.64	0.63	0.63	0.63
72	0.74	0.66	0.63	0.62	0.61	0.60	0.60	0.60
71	0.71	0.63	0.60	0.59	0.58	0.57	0.57	0.57
70	0.68	0.60	0.57	0.56	0.55	0.55	0.54	0.54
69	0.65	0.57	0.54	0.53	0.52	0.52	0.51	0.51
68	0.62	0.54	0.51	0.50	0.49	0.49	0.48	0.48
67	0.59	0.51	0.47	0.47	0.46	0.46	0.46	0.45
66	0.56	0.48	0.45	0.44	0.44	0.43	0.43	0.43
65	0.52	0.45	0.43	0.41	0.41	0.40	0.40	0.40
64	0.49	0.42	0.40	0.39	0.38	0.38	0.37	0.37
63	0.46	0.39	0.37	0.36	0.35	0.35	0.35	0.34
62	0.43	0.36	0.34	0.33	0.32	0.32	0.32	0.32
61	0.39	0.33	0.31	0.30	0.30	0.29	0.29	0.29
60	0.36	0.30	0.28	0.27	0.27	0.27	0.26	0.26
59	0.32	0.27	0.25	0.25	0.24	0.24	0.24	0.24
58	0.29	0.24	0.23	0.22	0.21	0.21	0.21	0.21
57	0.25	0.21	0.20	0.19	0.19	0.19	0.18	0.18
56	0.22	0.18	0.17	0.16	0.16	0.16	0.16	0.16
55	0.18	0.15	0.14	0.14	0.13	0.13	0.13	0.13
54	0.14	0.12	0.11	0.11	0.11	0.11	0.10	0.10
53	0.11	0.09	0.08	0.08	0.08	0.08	0.08	0.08
52	0.07	0.06	0.06	0.05	0.05	0.05	0.05	0.05
51	0.04	0.03	0.03	0.03	0.03	0.03	0.03	0.03
50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

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

Table 5. Quality Index Values for Estimating PWL (*continued*)

PWL	$n = 12$ to 14	$n = 15$ to 18	$n = 19$ to 25	$n = 26$ to 37	$n = 38$ to 69	$n = 70$ to 200	$n = 201$ to ∞
100	2.83	3.03	3.20	3.38	3.54	3.70	3.83
99	2.09	2.14	2.18	2.22	2.26	2.29	2.31
98	1.91	1.93	1.96	1.99	2.01	2.03	2.05
97	1.77	1.79	1.81	1.83	1.85	1.86	1.87
96	1.67	1.68	1.70	1.71	1.73	1.74	1.75
95	1.58	1.59	1.61	1.62	1.63	1.63	1.64
94	1.50	1.51	1.52	1.53	1.54	1.55	1.55
93	1.44	1.44	1.45	1.46	1.46	1.47	1.47
92	1.37	1.38	1.39	1.39	1.40	1.40	1.40
91	1.32	1.32	1.33	1.33	1.33	1.34	1.34
90	1.26	1.27	1.27	1.27	1.28	1.28	1.28
89	1.21	1.22	1.22	1.22	1.22	1.22	1.23
88	1.17	1.17	1.17	1.17	1.17	1.17	1.17
87	1.12	1.12	1.12	1.12	1.12	1.13	1.13
86	1.08	1.08	1.08	1.08	1.08	1.08	1.08
85	1.04	1.04	1.04	1.04	1.04	1.04	1.04
84	1.00	1.00	1.00	1.00	0.99	0.99	0.99
83	0.96	0.96	0.96	0.96	0.95	0.95	0.95
82	0.92	0.92	0.92	0.92	0.92	0.92	0.92
81	0.89	0.88	0.88	0.88	0.88	0.88	0.88
80	0.85	0.85	0.85	0.84	0.84	0.84	0.84
79	0.82	0.81	0.81	0.81	0.81	0.81	0.81
78	0.78	0.78	0.78	0.78	0.77	0.77	0.77
77	0.75	0.75	0.75	0.74	0.74	0.74	0.74
76	0.72	0.71	0.71	0.71	0.71	0.71	0.71
75	0.69	0.68	0.68	0.68	0.68	0.68	0.67
74	0.66	0.65	0.65	0.65	0.65	0.64	0.64
73	0.62	0.62	0.62	0.62	0.62	0.61	0.61
72	0.59	0.59	0.59	0.59	0.59	0.58	0.58
71	0.57	0.56	0.56	0.56	0.56	0.55	0.55
70	0.54	0.53	0.53	0.53	0.53	0.53	0.52
69	0.51	0.50	0.50	0.50	0.50	0.50	0.50
68	0.48	0.48	0.47	0.47	0.47	0.47	0.47
67	0.45	0.45	0.45	0.44	0.44	0.44	0.44
66	0.42	0.42	0.42	0.42	0.41	0.41	0.41
65	0.40	0.39	0.39	0.39	0.39	0.39	0.39
64	0.37	0.36	0.36	0.36	0.36	0.36	0.36
63	0.34	0.34	0.34	0.34	0.33	0.33	0.33
62	0.31	0.31	0.31	0.31	0.31	0.31	0.31
61	0.29	0.29	0.28	0.28	0.28	0.28	0.28
60	0.26	0.26	0.26	0.26	0.26	0.25	0.25
59	0.23	0.23	0.23	0.23	0.23	0.23	0.23
58	0.21	0.21	0.20	0.20	0.20	0.20	0.20
57	0.18	0.18	0.18	0.18	0.18	0.18	0.18
56	0.16	0.15	0.15	0.15	0.15	0.15	0.15
55	0.13	0.13	0.13	0.13	0.13	0.13	0.13
54	0.10	0.10	0.10	0.10	0.10	0.10	0.10
53	0.08	0.08	0.08	0.08	0.08	0.08	0.08
52	0.05	0.05	0.05	0.05	0.05	0.05	0.05
51	0.03	0.03	0.03	0.03	0.03	0.03	0.02
50	0.00	0.00	0.00	0.00	0.00	0.00	0.00

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

Table 6. A PWL Estimation Table for Sample Size *n* = 5

Q	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	50.00	50.36	50.71	51.07	51.42	51.78	52.13	52.49	52.85	53.20
0.1	53.56	53.91	54.27	54.62	54.98	55.33	55.69	56.04	56.39	56.75
0.2	57.10	57.46	57.81	58.16	58.52	58.87	59.22	59.57	59.92	60.28
0.3	60.63	60.98	61.33	61.68	62.03	62.38	62.72	63.07	63.42	63.77
0.4	64.12	64.46	64.81	65.15	65.50	65.84	66.19	66.53	66.87	67.22
0.5	67.56	67.90	68.24	68.58	68.92	69.26	69.60	69.94	70.27	70.61
0.6	70.95	71.28	71.61	71.95	72.28	72.61	72.94	73.27	73.60	73.93
0.7	74.26	74.59	74.91	75.24	75.56	75.89	76.21	76.53	76.85	77.17
0.8	77.49	77.81	78.13	78.44	78.76	79.07	79.38	79.69	80.00	80.31
0.9	80.62	80.93	81.23	81.54	81.84	82.14	82.45	82.74	83.04	83.34
1.0	83.64	83.93	84.22	84.52	84.81	85.09	85.38	85.67	85.95	86.24
1.1	86.52	86.80	87.07	87.35	87.63	87.90	88.17	88.44	88.71	88.98
1.2	89.24	89.50	89.77	90.03	90.28	90.54	90.79	91.04	91.29	91.54
1.3	91.79	92.03	92.27	92.51	92.75	92.98	93.21	93.44	93.67	93.90
1.4	94.12	94.34	94.56	94.77	94.98	95.19	95.40	95.61	95.81	96.01
1.5	96.20	96.39	96.58	96.77	96.95	97.13	97.31	97.48	97.65	97.81
1.6	97.97	98.13	98.28	98.43	98.58	98.72	98.85	98.98	99.11	99.23
1.7	99.34	99.45	99.55	99.64	99.73	99.81	99.88	99.94	99.98	100.00

Notes:

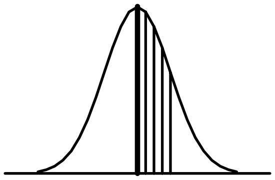
Values in the body of the table are estimates of PWL corresponding to specific values of $Q_L = (\overline{X} - L)/s$ or $Q_U = (U - \overline{X})/s$. For negative *Q* values, the table values must be subtracted from 100.

A more complete set of PWL tables in this format, for sample sizes from *n* = 3 to *n* = 30, is available. (18)

Table 7. Another PWL Estimation Table for Sample Size *n* = 5

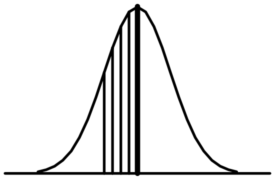
Q_L or Q_U	PWL_L or PWL_U
1.671 or More	100
1.601 to 1.670	99
1.541 to 1.600	98
1.491 to 1.540	97
1.441 to 1.490	96
1.391 to 1.440	95
1.351 to 1.390	94
1.311 to 1.350	93
1.271 to 1.310	92
1.231 to 1.270	91
1.191 to 1.230	90
1.151 to 1.190	89
1.121 to 1.150	88
1.081 to 1.120	87
1.051 to 1.080	86
1.011 to 1.050	85
0.981 to 1.010	84
0.951 to 0.980	83
0.911 to 0.950	82
0.881 to 0.910	81
0.851 to 0.880	80
0.821 to 0.850	79
0.781 to 0.820	78
0.751 to 0.780	77
0.721 to 0.750	76
0.691 to 0.720	75
0.661 to 0.690	74
0.631 to 0.660	73
0.601 to 0.630	72
0.571 to 0.600	71
0.541 to 0.570	70
0.511 to 0.540	69
0.471 to 0.510	68
0.451 to 0.470	67
0.431 to 0.450	66
0.401 to 0.430	65
0.371 to 0.400	64
0.341 to 0.370	63
0.311 to 0.340	62
0.281 to 0.310	61
0.251 to 0.280	60
0.231 to 0.250	59
0.201 to 0.230	58
0.171 to 0.200	57
0.141 to 0.170	56
0.111 to 0.140	55
0.081 to 0.110	54
0.061 to 0.080	53
0.031 to 0.060	52
0.001 to 0.030	51
-0.029 to 0.000	50
-0.029 to 0.000	50
-0.059 to -0.030	49
-0.079 to -0.060	48
-0.109 to -0.080	47
-0.139 to -0.110	46
-0.169 to -0.140	45
-0.199 to -0.170	44
-0.229 to -0.200	43
-0.249 to -0.230	42
-0.279 to -0.250	41
-0.309 to -0.280	40
-0.339 to -0.310	39
-0.369 to -0.340	38
-0.399 to -0.370	37
-0.429 to -0.400	36
-0.449 to -0.430	35
-0.469 to -0.450	34
-0.509 to -0.470	33
-0.539 to -0.510	32
-0.569 to -0.540	31
-0.599 to -0.570	30
-0.629 to -0.600	29
-0.659 to -0.630	28
-0.689 to -0.660	27
-0.719 to -0.690	26
-0.749 to -0.720	25
-0.779 to -0.750	24
-0.819 to -0.780	23
-0.849 to -0.820	22
-0.879 to -0.850	21
-0.909 to -0.880	20
-0.949 to -0.910	19
-0.979 to -0.950	18
-1.009 to -0.980	17
-1.049 to -1.010	16
-1.079 to -1.050	15
-1.119 to -1.080	14
-1.149 to -1.120	13
-1.189 to -1.150	12
-1.229 to -1.190	11
-1.269 to -1.230	10
-1.309 to -1.270	9
-1.349 to -1.310	8
-1.389 to -1.350	7
-1.439 to -1.390	6
-1.489 to -1.440	5
-1.539 to -1.490	4
-1.599 to -1.540	3
-1.669 to -1.600	2
-1.789 to -1.670	1
-1.790 or Less	0

Table 8. Areas Under the Standard Normal Distribution



Z

or



–Z

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3183
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998

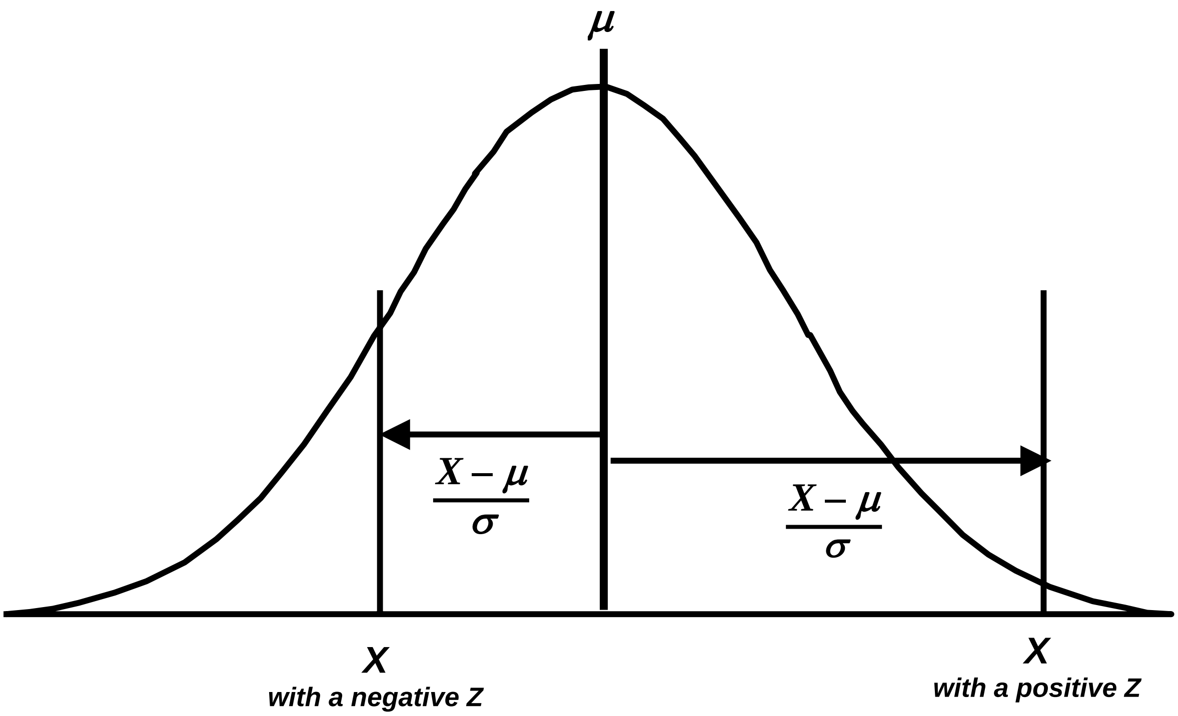


Figure 12. Illustration of the Calculation of the *Z*–statistic

Conceptually, the *Q*–statistic, or *quality index*, performs identically the same function as the *Z*–statistic except that now the reference point is the mean of an individual sample, $\overline{X}$, instead of the population mean, μ, and the points of interest with regard to areas under the curve are the specification limits.

$$Q_L = \frac{\overline{X} - LSL}{s} \tag{11}$$

and

$$Q_U = \frac{USL - \overline{X}}{s} \tag{12}$$

where: Q_L = quality index for the lower specification limit.
Q_U = quality index for the upper specification limit.
LSL = lower specification limit.
USL = upper specification limit.
$\overline{X}$ = the sample mean for the lot.
s = the sample standard deviation for the lot.

The Q–statistic, therefore, represents the distance in sample standard deviation units that the sample mean is offset from the specification limit. A positive Q–statistic represents the number of sample standard deviation units that the sample mean falls *inside* the specification limit. Conversely, a negative Q–statistic represents the number of sample standard deviation units that the sample mean falls *outside* the specification limit. These cases are illustrated in figures 13 and 14.

Q_L is used when there is a one–sided lower specification limit, while Q_U is used when there is a one–sided upper specification limit. For two–sided specification limits, the PWL value is estimated as:

$$\mathrm{PWL}_T = \mathrm{PWL}_U + \mathrm{PWL}_L - 100 \tag{13}$$

where:
PWL_U = percent below the upper specification limit (based on Q_U).
PWL_L = percent above the lower specification limit (based on Q_L).
PWL_T = percent within the upper and lower specification limits.

Example

An example using a simplified portland cement concrete specification can be used to explain the PWL concept. In this example, the minimum specification limit for strength is 21,000 kPa. One requirement of the PWL procedure is that the sample size must be greater than $n = 2$ since both the sample mean and sample standard deviation are necessary to estimate PWL. For this specification, the sample size has been chosen as $n = 4$. Furthermore, the specification requires that at least 95 percent of the lot exceed the minimum strength (i.e., PWL ≥ 95). Table 5 shows that the minimum Q value is 1.35 for 95 PWL and a sample size of $n = 4$. Whenever the mean is $1.35s$ above the specification limit, the lot is accepted. However, as used most frequently, the Q value will be used to compute the PWL and that will, in turn, be used to determine a payment factor.

For example, suppose that the acceptance tests for a lot have a sample mean of 25,000 kPa and a sample standard deviation of 3,400 kPa. Does this lot meet the specification requirement? The quality index value is calculated as:

$$Q_L = \frac{25000 - 21000}{3400} = 1.18 \tag{14}$$

Using this Q–value, $n = 4$, and table 5, the estimated PWL for the lot is between 89 and 90. This is less than the required 95, so the lot does not meet the specified strength requirement.

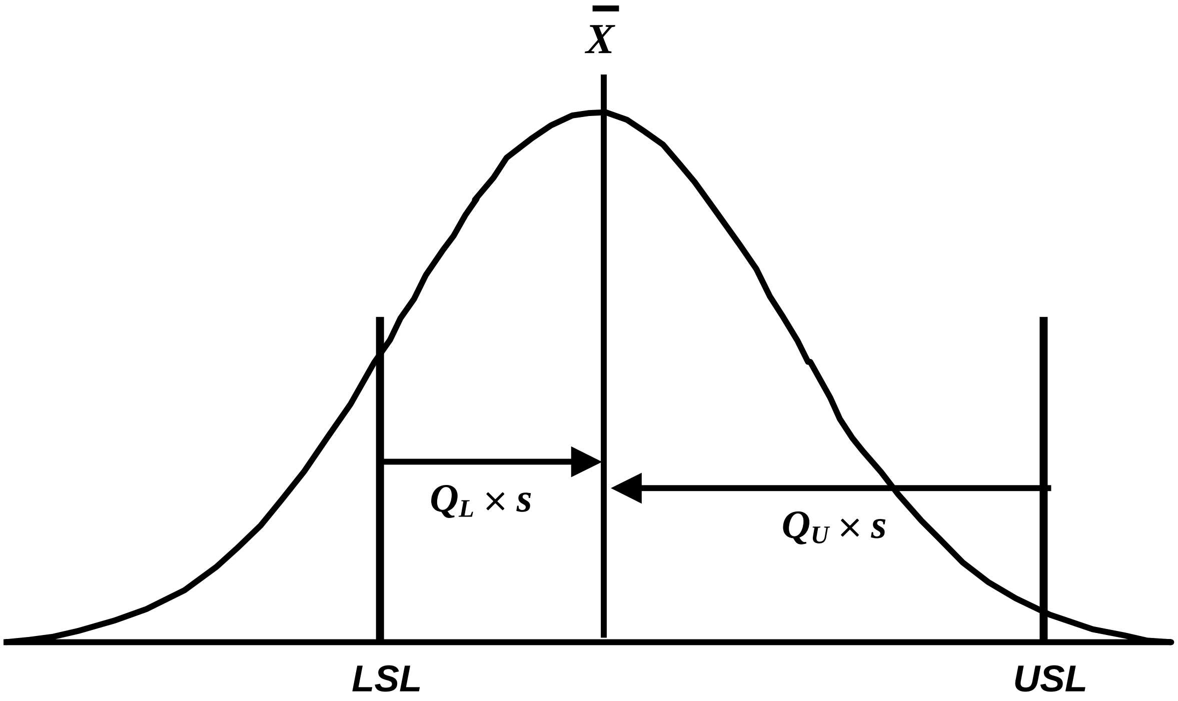


Figure 13. Illustration of Positive Quality Index Values

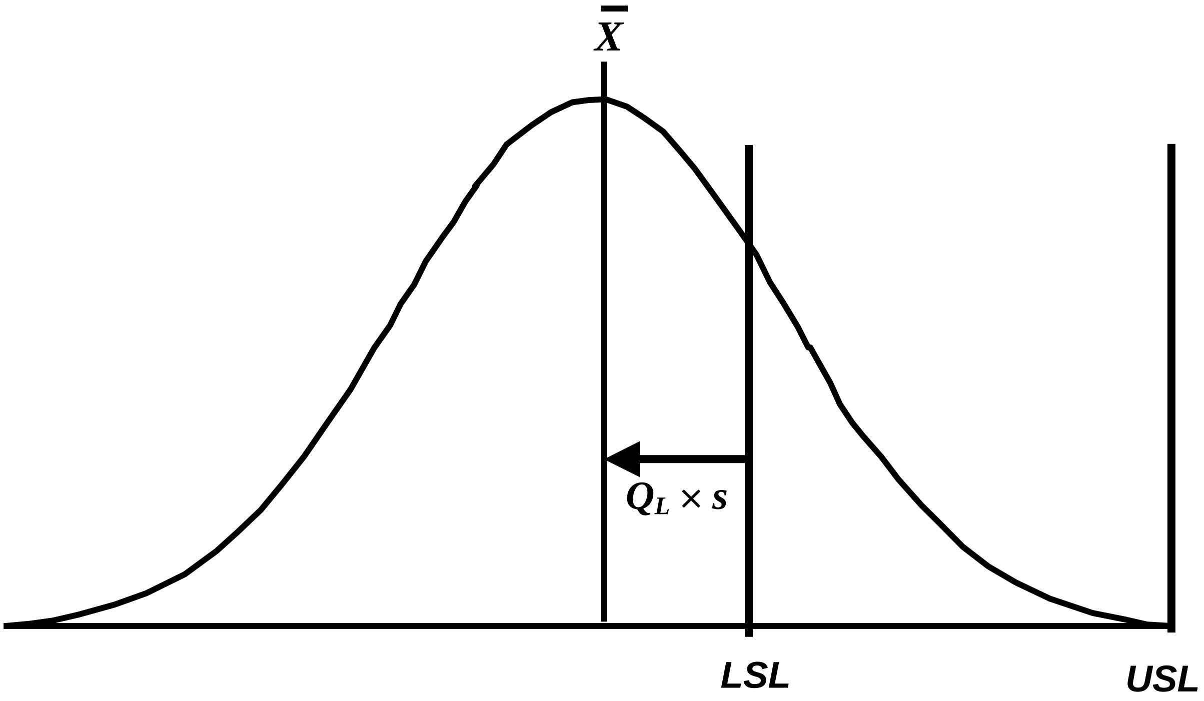


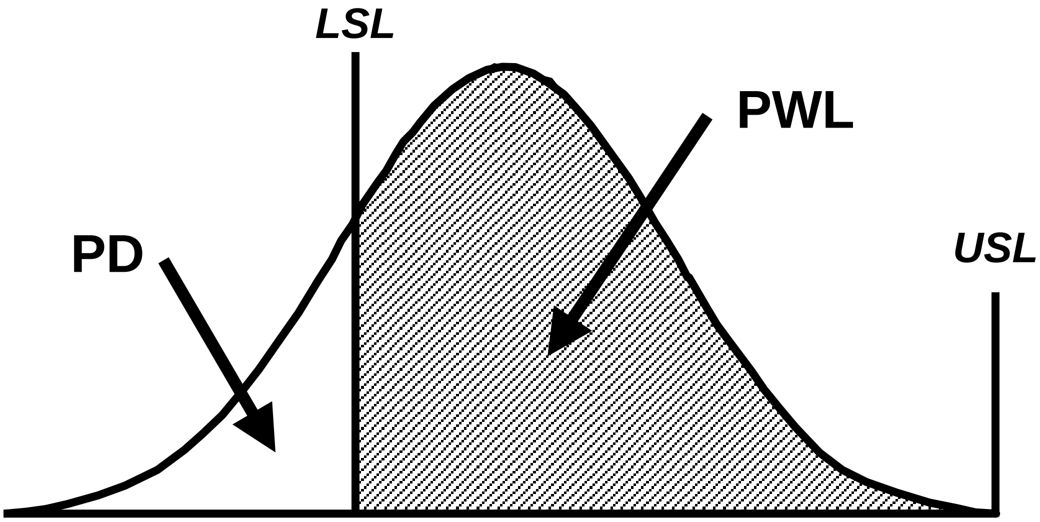
Figure 14. Illustration of a Negative Quality Index Value

Intuitively, PWL is a good measure of quality since it is reasonable to assume that the more of the material that is within the specification limits, the better the quality of the product should be. A detailed discussion and analysis of the PWL measure of quality is presented in the technical report for the project. (17)

32.2. The PD Quality Measure

As noted in the above definition for PWL, PD is related to the PWL by the simple relationship PWL = 100 – PD. The use of PD as a quality measure can have some advantages, particularly with two–sided specifications, because the PD below the lower specification limit can simply be added to the PD above the upper specification limit to obtain the total PD value. This is slightly easier than equation 13 that is required when using PWL. The relationship between PD and PWL is shown in figure 15.

Since PD and PWL can be converted to one another simply by subtracting from 100, they are equivalent quality measures. Therefore, any discussion of PWL will apply equally to PD. Most agencies have preferred to measure quality in terms of how much of the material meets the requirements, i.e., PWL, rather than to measure how much does not meet the requirements, i.e., PD. This approach to measuring how much is "good" as opposed to how much is "bad" seems to have been more popular among agencies. The PWL approach, rather than PD, has also been promoted by the FHWA. For these reasons, most discussions in this manual center on PWL rather than PD. However, examples and case studies for both methods are included.



15. Relationship between PWL and PD

32.3. The AAD Quality Measure

For specifications that have a target value, the average deviation from the target value has in the past sometimes been used as a means for determining acceptability of the product. This approach can have the effect of encouraging the contractor to manipulate its process during the production of a lot. For example, if two test results in the morning are below the target value, there is a strong incentive for the contractor to increase the process mean in the afternoon in an effort to get two higher test results so that the average of the four tests for the lot will be near the target value. In essence, this acceptance approach encourages the contractor to increase process variability by making frequent adjustments to the process mean.

Recommendation

The **average deviation** from the target value **should NOT be used** as the quality measure for QA acceptance plans.

To avoid the problem of over–adjusting the process in response to early test results, the average absolute deviation from the target can be used for the acceptance decision. The TRB glossary (2) includes the following definition:

- ***AAD****—For a series of test results, the mean of absolute deviations from a target or specified value. [A low AAD implies both good accuracy and good precision; a high AAD, however, does not necessarily imply both poor accuracy and poor precision (i.e., accuracy or precision, but not both, might be quite good).]*

The equation for calculating AAD is as follows:

$$AAD = \frac{\sum |X_i - T|}{n} \qquad (15)$$

where: X_i = individual test results.
T = target value.
n = number of tests per lot.

By taking the absolute value of the deviation from the target, the contractor cannot benefit by any strategy other than aiming for the target value. For example, if two early tests on the lot indicate values below the target by –0.4 and –0.6, using AAD the contractor cannot offset these low values with two later values of +0.5 and +0.5. This is true because, while the average of these four numbers is (–0.4 –0.6 +0.5 +0.5) ÷ 4 = 0.0, the average of the absolute values of these numbers is (0.4 + 0.6 + 0.5 +0.5) ÷ 4 = 0.5.

Intuitively, AAD is a good measure of quality since it is reasonable to assume that the lower the average absolute deviation, the closer the process is to the target and hence the better the quality of the product should be. There are, however, some disadvantages associated with AAD acceptance plans. One primary drawback is that they can only be used when there is a target value. They cannot, therefore, be used when there is only one specification limit, such as might be the case with concrete compressive strength.

Another drawback with using AAD is that the variability of the material in the lot may not be adequately measured. Specifically, very different sets of test results can give identical AAD values. For example, the three sets of four tests shown in table 9 each give the same value for AAD. Each set of tests would therefore yield the same payment for a lot. Not only must it be wondered if equal payment is appropriate for these widely different conditions, the use of AAD fails to document these differences that could be useful for considering future modifications of the specification. Specifically, the sample means and sample standard deviations (measure of variability) vary considerably for the three cases. With acceptance based solely on AAD, these mean and variability differences would not be identified.

A detailed discussion and analysis of the AAD measure of quality is presented in the technical report for the project. (17)

Table 9. Example of Lots with Similar AAD Values but Different Within–Lot Means and Standard Deviations

	X – T		
Test	**Lot 1**	**Lot 2**	**Lot 3**
1	+0.4	+0.4	–0.4
2	–0.5	+0.5	–0.5
3	+0.6	+0.6	–0.6
4	–0.5	+0.5	–0.5
AAD	0.5	0.5	0.5
Sample Mean	0.0	+0.5	–0.5
Sample Std Dev	0.58	0.08	0.08

32.4. The CI Quality Measure

The TRB glossary (2) includes the following definition:

- ***CI****—A measure of the dispersion of a series of results around a target or specified value, expressed as the square root of the quantity obtained by summing the squares of the deviations from the target value and dividing by the number of observations.*

  The standard deviation is a measure of precision, but the CI is a measure of exactness (accuracy) or degree of conformance with the target.

Conceptually, the CI is similar to the AAD. While the AAD uses the average of the absolute values of the individual deviations from the target value, the CI uses the squares of the individual deviations from the target value. The CI is also similar in concept to the standard deviation. The standard deviation is the root mean square of differences from the mean; whereas the CI is the root mean square of differences from a target such as the job mix formula for HMAC. Like the AAD, the CI discourages mid–lot process adjustments by not allowing positive and negative deviations from the target to cancel out one another. As shown in the TRB glossary definition above, the CI is calculated as follows:

$$CI = \sqrt{\frac{\sum (X_i - T)^2}{n}} \tag{16}$$

where: X_i = individual test results.
T = target value.
n = number of tests per lot.

The CI is very similar in practice to the AAD, and has the same disadvantages of not being appropriate for a one–sided specification and of potentially having the same CI value for vastly different sample results. Table 10 shows the CI, sample mean, and sample standard deviation values for three example lots.

A detailed discussion and analysis of the CI measure of quality is presented in the technical report for the project. (17)

Table 10. Example of Lots with Similar CI Values but Different Within–Lot Means and Standard Deviations

Test	Lot 1	Lot 2	Lot 3
1	+1.4	+1.4	–1.4
2	–1.5	+1.5	–1.5
3	+1.6	+1.6	–1.6
4	–1.5	+1.5	–1.5
CI	1.50	1.50	1.50
Sample Mean	0.0	+1.5	–1.5
Sample Std Dev	1.73	0.08	0.08

32.5. Moving Average

A few agencies have developed acceptance procedures based on the moving average of the quality characteristic. For moving averages first a sample size, say *n* = 4, must be determined. The first average is then determined from the first four values. For the second moving average, the fifth value replaces the first value in the calculations. For the third moving average, the sixth value replaces the second value, and so on. This process is illustrated in figure 16.

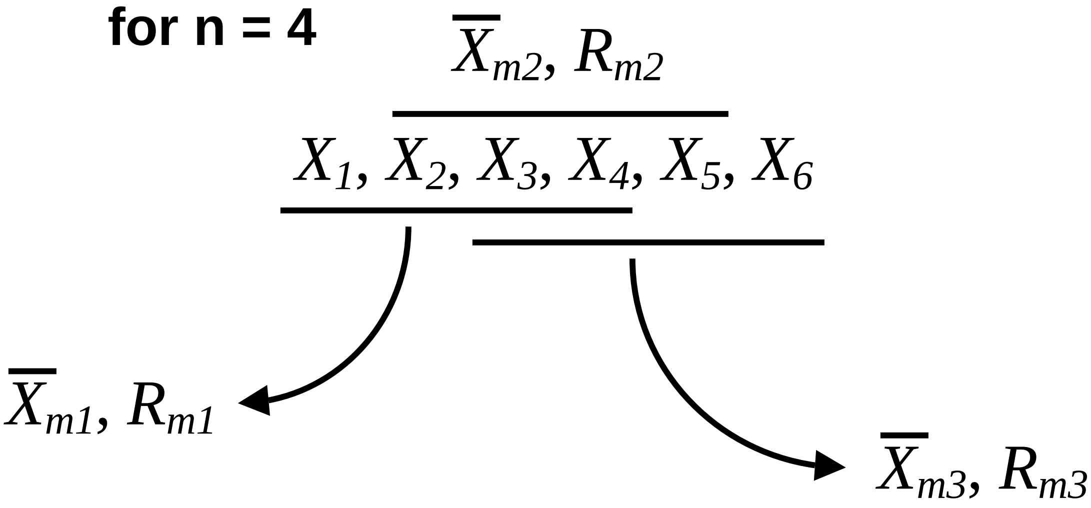


Figure 16. Illustration of the Moving Average

The use of the moving average provides a smoothing effect compared to plotting individual test results. Moving averages have typically been applied for process control purposes, and are particularly appropriate when continuous processes are involved. Moving averages have frequently been used for process control in the manufacture of chemicals.

Moving averages were not developed for use as an acceptance approach, and they have several drawbacks when applied in acceptance plans. Due to the nature in which it is calculated, the use of the moving average is not consistent with the use of lot–by–lot acceptance. Since each individual test result appears in multiple moving averages, it is difficult to determine when or where a lot begins or ends. Also, when acceptance is based on a lot, it is assumed that the various lots are independent of one another and that the material within each lot represents material from a constant process or population. Since each individual test result appears in several moving averages, the moving averages are obviously correlated and the results of one average are not independent of the next.

The use of moving averages also does not lend itself well to determining payment factors. Since successive moving averages are correlated, and since individual lots are not well defined, it is not easy to determine payment factors on a lot–by–basis. As a result, acceptance procedures based on the moving average often result in production shut downs and plant adjustments rather than determining appropriate payment factors for specific production lots.

Recommendation

The **moving average should NOT be used** as the quality measure for QA acceptance plans.

32.6. Recommended Quality Measure

It is necessary to measure both center and spread when characterizing a lot (population) of material. Even if appropriate typical process standard deviation and target miss values are used to establish the acceptance procedures, there are potential difficulties with AAD and CI quality measures. One drawback with AAD and CI acceptance plans is that since lot variabilities are not directly measured, a given lot AAD or CI can come from a number of different populations. For example, the population could be centered at the target, but have a relatively large standard deviation, i.e., larger than the one that was assumed when developing the AAD or CI acceptance limits or payment equation. Another population could have the same AAD or CI by being centered off the target and having the same standard deviation that was assumed when developing the acceptance values. A third population could have the same AAD or CI value by having a mean far from the target, but also having a relatively small standard deviation.

While some of these drawbacks may also apply to PWL (or PD) acceptance plans, such as the fact that a given PWL can represent many different populations, there are fewer drawbacks due to the fact that both sample mean and standard deviation are determined in the PWL method. Also, since the PWL method can be used with both one–sided and two–sided acceptance properties, it is more versatile since it does not require separate approaches for one–sided and two–sided cases. The PWL approach has been endorsed and recommended by the FHWA for many years, and it is also the method used in the *AASHTO QA Guide Specification*.(7)

Recommendation

PWL is the recommended quality measure. It should be noted that PD is equally suitable and provides all the same mathematical advantages as PWL. The case studies in chapter 9, for example, use PD as the quality measure. A detailed analysis and evaluation of the PWL, PD, AAD, and CI measures, including how non–normal populations impact these measures, is included in the technical report for this project.(17) Any agency that is considering using the AAD or CI quality measures should thoroughly review the technical report before making a final decision.

33 Screening Test Quality Characteristics: Determine Quality Measure to Use

This is similar to the decision for quality measures that was made for payment determination characteristics, with the possible exception of the number of test results available for the decision. Because there may be only a few test results for the pass/fail decision, usually one or two, this may be insufficient to allow a rigorous statistical analysis. Thus, the acceptance decision may have to be made on a single test or the average of two tests. As mentioned above, this will increase the risks involved. Due to the small sample sizes involved, PWL and PD quality measures are not likely to be options. In most cases, the screening test becomes a simple pass/fail attributes plan. That is, does the one test fall within the allowable limits or does it not?

34 Payment Quality Characteristics: Determine AQL, RQL, and Specification Limits

The AQL and RQL must be defined, and the specification limits and acceptance limits must be determined. The AQL, RQL, and specification limits, are intimately related and the decisions regarding these are typically made concurrently. Decisions regarding acceptance limits are related to the risks to the contractor and agency, and the acceptance limits are typically determined based on a risk analysis (see chapter 7).

34.1. Definitions

Specification and acceptance limits are often confused. The TRB glossary (2) provides the following definitions:

- ***Specification limit(s)****—The limiting value(s) placed on a quality characteristic, established preferably by statistical analysis, for evaluating material or construction within the specification requirements. The term can refer to either an individual upper or lower specification limit, USL or LSL, called a single specification limit; or to USL and LSL together, called double specification limits.*

- ***Acceptance limit—****In variables acceptance plans, the limiting upper or lower value, placed on a quality measure, that will permit acceptance of a lot. [Unlike specification limits placed on a quality characteristic, an acceptance limit is placed on a quality measure. For example, in PWL acceptance plans, PWL refers to specification limits placed on the quality characteristic, and the minimum allowable PWL identifies the acceptance limit for the PWL quality measure.]*

Specification limits are based on engineering requirements and are usually expressed in the same units as those used for the quality characteristic of concern. The acceptance limits are usually expressed in statistical units (e.g., mean, PD, PWL, AAD, etc.). For accept or reject acceptance plans, the acceptance limits are established from a risk analysis (see chapter 7). For acceptance plans with payment adjustment provisions, additional acceptance limits, often expressed in the form of an equation or equations, are used to distinguish among the various possible payment levels. Payment provisions are presented later in this chapter, while the analysis of risks associated with payment adjustment provisions is presented in chapter 7.

The TRB glossary (2) offers the following definitions for AQL and RQL:

- ***AQL****—That minimum level of actual quality at which the material or construction can be considered fully acceptable (for that quality characteristic). For example, when quality is based on PWL, the AQL is that actual (not estimated) PWL at which the quality characteristic can just be considered fully acceptable. [Acceptance plans should be designed so that AQL material will receive an EP of 100 percent.]*

- ***RQL****—That maximum level of actual quality at which the material or construction can be considered unacceptable (rejectable). For example, when quality is based on PD, the RQL is that actual (not estimated) PD at which the quality characteristic can just be considered fully rejectable. [It is desired to require removal and replacement, corrective action, or the assignment of a relatively low pay factor when RQL work is detected.]*

34.2. Setting the Limits

Establishing specification requires defining acceptable (AQL) and unacceptable (RQL) material. These are both engineering decisions. The AQL decision defines acceptable material, and should address the material that will provide satisfactory service at an affordable cost when used for the intended purpose. What constitutes acceptable material is often determined based on what has performed well in the past. However, it is preferable if performance data are available to quantify performance. Statistics has been a valuable tool in defining the parameters (mean and standard deviation) for acceptable material. Caution should be exercised if a lower variability is chosen for the specification than has been shown to be readily achievable. Arbitrarily "tightening the specs" can increase the cost of the material above that which may be cost effective, or even attainable.

In addition to defining acceptable material, a sometimes more difficult decision must be made regarding what constitutes unacceptable, or RQL, material. Unacceptable material is that which is unlikely to perform as planned. It should have a low probability of being accepted or will be accepted only under the conditions of a reduced payment schedule.

Selecting the specification limit(s) must be done in concert with the quality measure and the AQL and RQL. For instance, if PWL is used, the AQL might be set at 90 PWL. This means that when the sample statistics estimate the population to have 90 percent of the product within specification limits, the product is completely acceptable. However, it is conceivable that comparable product could be defined at an AQL of 85 PWL with more restrictive specification limits, or many other possible combinations of AQL and specification limits. An example may help to clarify this situation.

Example

Suppose that an agency has decided, based on a large amount of project data that it collected and analyzed, that a “typical” standard deviation (see chapter 5) for a lot defined in the acceptance plan for asphalt content for HMAC is 0.18 percent. How could this information be used to establish specification limits, and AQL and RQL values for asphalt content?

Decide on Quality Measure. The first issue to address is what measure of quality is to be used. Assume that PWL will be used as the quality measure. How can this measure be related to the definitions of AQL and RQL materials and to the specification limits?

Define AQL Material. Since asphalt content has a stipulated target value, i.e., the job mix formula (JMF) asphalt content, the agency may choose to define AQL material as a lot for which the average asphalt content is equal to the JMF target value and for which the standard deviation is equal to or less than the “typical” value of 0.18 percent. This defines AQL material in terms of the desired population mean and standard deviation, but the AQL definition must also be related to the required quality measure, which in this case is PWL.

Set Specification Limits. The specification limits and the AQL are related. For example, the agency might decide to set the AQL as a value of 90 PWL. This selection of 90 PWL for the AQL is arbitrary, but is a commonly used value. The AQL population defined from past projects in terms of mean and standard deviation should just meet this PWL definition for AQL. So, in this case, the specification limits would be set such that a population with a mean at the JMF and a standard deviation of 0.18 percent would have 90 percent of its area within the specification limits. These limits can be determined by finding the Z–value from a standard normal table that corresponds to an area of 0.90 within the mean $\pm Z$ standard deviations (i.e., $\mu \pm Z\sigma$).

Table 11 presents some typical $\pm Z$ regions within which selected areas of the normal distribution fall. From this table it is seen that 0.90 (or 90 percent) of the normal distribution falls within ±1.645 standard deviations from the population mean. Figure 17 shows a graphical representation of a population of this AQL material.

The specifications might therefore be set at the JMF asphalt content plus or minus 1.645 times the typical standard deviation value, or JMF ± 1.645(0.18 percent) = JMF ± 0.30 percent. In this case, the AQL is 90 PWL and the specification limits are JMF ± 0.30 percent.

Alternatively, using the same AQL population, the agency could decide to establish the AQL as 85 PWL. In this case, the specification limits would be set at the JMF plus or minus 1.439 times the typical standard deviation value (see table 11), or JMF ± 1.439(0.18 percent) = JMF ± 0.26 percent. The specification limits are different in this case because the definition for AQL in terms of PWL is different.

Define RQL Material. The RQL must now be defined. There is no single correct way to establish either the AQL or the RQL. In this case, once the AQL and specification limits are established, the RQL could be established in a number of ways. One way would be to decide that the material should be rejected once a "large" percentage of material is outside the specification limits. What constitutes a "large" percentage would then need to be decided. The agency could decide that the material is "bad" once half of it is outside of the specification limits. In this case, the RQL would be established as a PWL value of 50. Any lot with an estimated PWL value of 50 or less would then be required to be removed and replaced.

Alternately, the highway agency might base the definition of RQL on the analysis of past project data. For instance, the agency might decide that past projects had performed inadequately when the average asphalt content was 0.25 percent above or below the JMF target value. In this case, the agency might decide to set the RQL based upon the PWL value for a population that has a mean 0.25 percent above or below the JMF target and that has a standard deviation equal to the "typical" value of 0.18 percent. The PWL for the RQL population depends upon which specification limits, those based on AQL = 90 PWL or those based on AQL = 85 PWL, apply. Figure 18 illustrates the case of the RQL population when the specification limits are JMF ± 0.30 percent. In this case, it can be seen that the PWL for the RQL population corresponds to the area of the population that lies within the specification limits. In this case, the RQL might be defined as a lot with a PWL value of 60.

This approach to defining RQL material contains a number of simplifying assumptions. For example, it looked only at how far the population mean departed from the target value and did not consider the population standard deviation. This, in essence, assumes that the typical standard deviation value of 0.18 percent will be achieved on all projects. This approach also does not consider the interaction and effect of other quality characteristics, such as density, thickness, etc., on the performance of past projects.

Table 11. $\mu \pm Z\sigma$ Regions for Selected Areas Under the Normal Distribution

Area	**0.99**	**0.95**	**0.90**	**0.85**	**0.80**
$\pm Z$	2.576	1.960	1.645	1.439	1.282

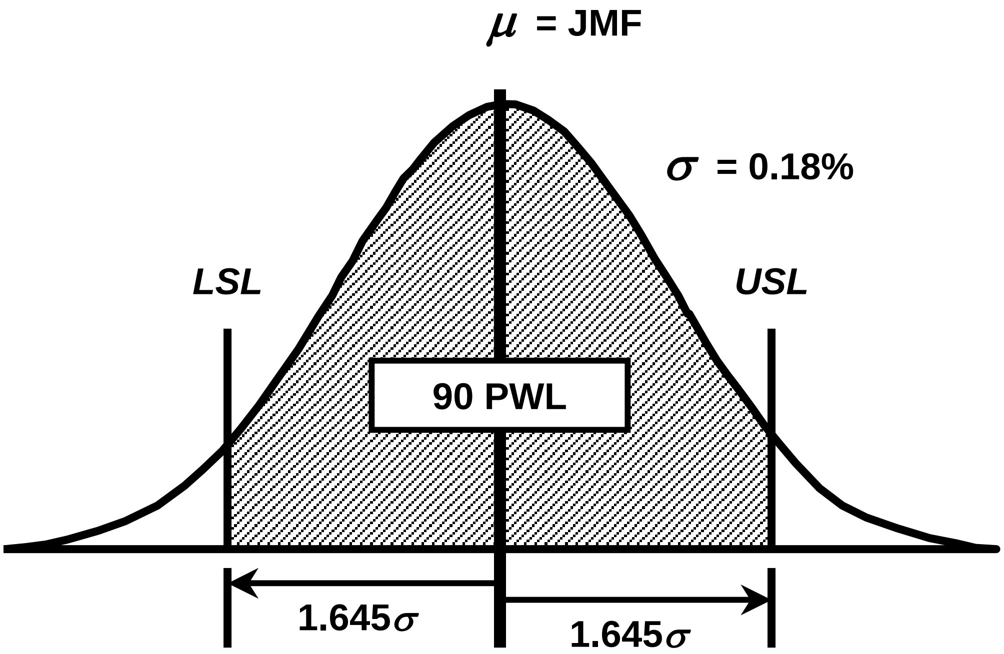


Figure 17. AQL Material

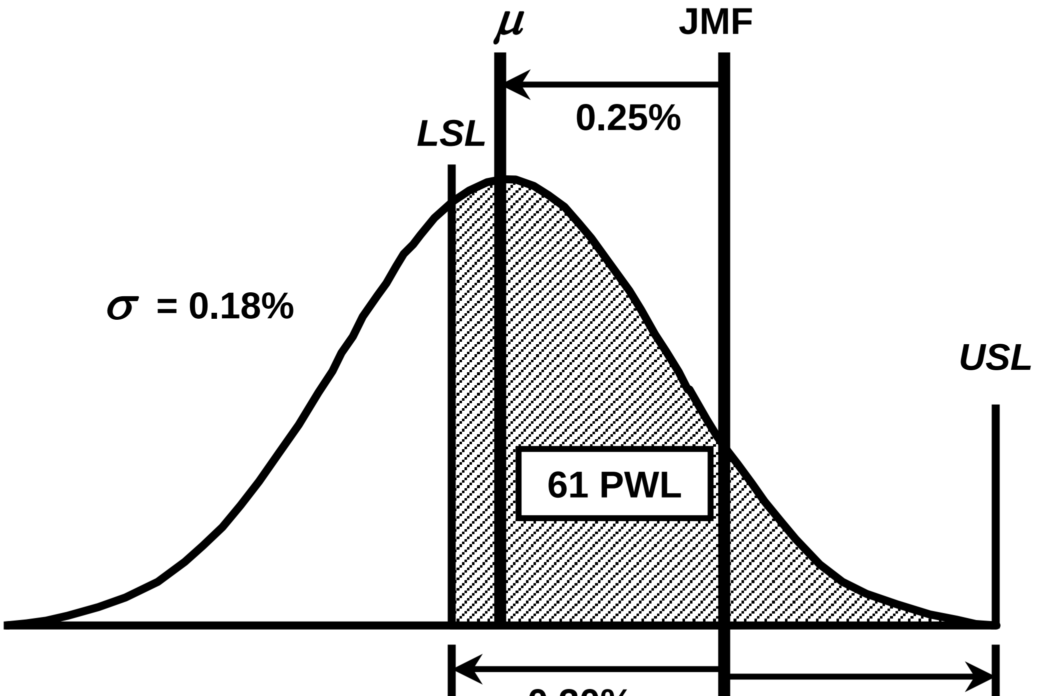


Figure 18. RQL Material

Ideally, since the current trend is to write PRS for which the quality measure is related to expected performance in some known, quantitative way, the agency may wish to analyze past project data in terms of the chosen quality measure. For example, if the agency has chosen PWL as the statistical quality measure, it may be worthwhile to use any available data (which might be in–house data, or equivalent data from other agencies) to seek a relationship between PWL and expected performance. If sufficient data are available, this may be the most direct way to determine realistic values for both the AQL and the RQL. Various methods to develop suitable performance relationships are discussed in more detail later in this manual.

As noted previously, there is no single "correct" method for establishing the AQL and RQL values and the specification limits. Another example may help to further illustrate how these are all related and how they can be established.

Example

An agency needs to establish an acceptance plan for the percent passing the 75 micron (μm) sieve for an aggregate base course. Experience from past projects indicates that base courses perform well if the amount of material passing the 75 μm sieve is 7 percent or less, and that they perform poorly if the amount of material passing the 75 μm sieve exceeds 10 percent. A typical standard deviation for this material has been found from analysis of past project data to be about 1.1 percent.

Decide on Quality Measure. From the past project information in the previous paragraph, the agency believes that the base will perform well as long as most of the material has less than 7 percent passing the 75 μm sieve. This indicates that a convenient quality measure is the percentage of the material with less than 7 percent passing the 75 μm sieve. This makes PWL a convenient and appropriate quality measure.

Define AQL Material. Based on the information in the preceding paragraph, to define the AQL it will be necessary for the agency to decide what PWL value corresponds to "most." While this is an arbitrary decision, the agency might choose to define "most" as 90 percent or more. Other choices are obviously possible, but 90 PWL is a common choice. Thus, the AQL is set at 90 PWL. This is a relatively conservative definition because, even if the standard deviation were considerably larger than the typical value, there is little chance that any of the material in the normal distribution representing AQL quality would reach the known critical value of 10 percent passing the 75 μm sieve. The diagram in figure 19 illustrates AQL material.

Define RQL Material. The PWL value for the RQL must also be determined. If the extreme upper tail of a normal distribution with a standard deviation of 1.1 percent is placed at the known critical value of 10 percent, then the mean of that distribution will be at approximately 10 percent – (3 × 1.1 percent) = 6.7 percent. The table of areas under the standard normal curve, table 8, can be used to determine that this corresponds to approximately 60 percent of the population below the known satisfactory value of 7 percent (or approximately 40 percent above the satisfactory value). On those occasions where the standard deviation was larger than the typical value of 1.1 percent, a relatively small portion of the distribution would extend above the critical value of 10 percent. As the amount of material with 7.0 percent or less passing the 75 μm sieve decreases below 60 percent, however, progressively more will exceed the critical value of 10 percent and serious performance problems might be expected to develop. Thus, the RQL is chosen as 60 PWL. The diagram in figure 20 illustrates RQL material.

Set Specification Limit. The upper specification limit, based on past project data and the definition of AQL, is 7.0 percent passing the 75 μm sieve.

Caution

Because of the severe consequences imposed when RQL work is detected, such as requiring removal and replacement at the contractor's expense, or the assignment of a minimum payment factor, it is important that the RQL be set at a sufficiently low level of quality that the agency could, if challenged, defend this decision. An additional reason for setting the RQL at a relatively low level of quality is to reduce the risk of mistakenly identifying an RQL condition due to the imprecision of the sampling and testing process. (See further discussion of this in chapter 7 on evaluating risks.)

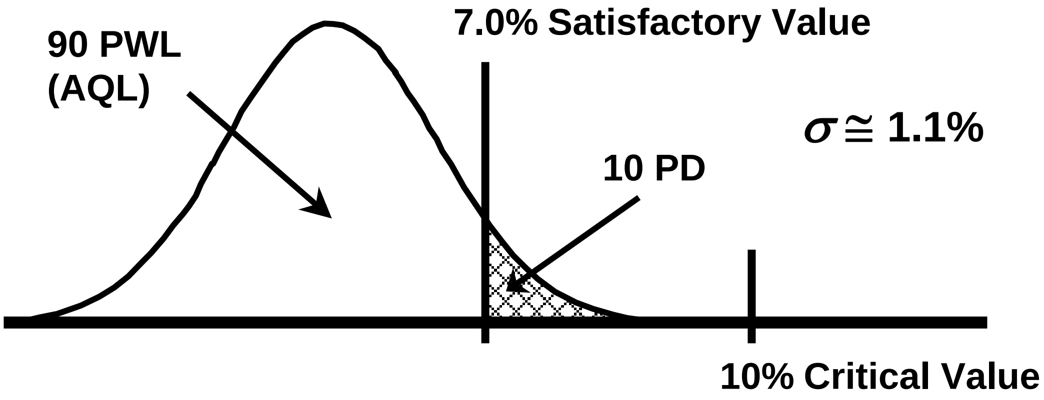


Figure 19. AQL Material

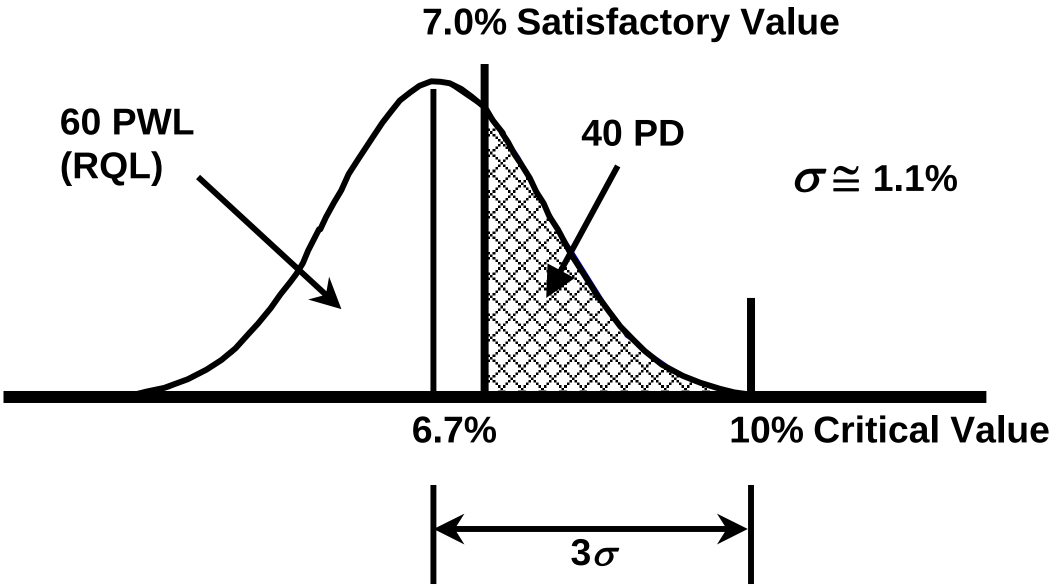


Figure 20. RQL Material

35 Screening Test Quality Characteristics: Determine AQL, RQL and Specification Limits

For screening tests, the determination of the specification limits is usually simpler than that for payment determination. The acceptance plan for screening quality characteristics may have specification limits, acceptance limits, or both. The specification limits are the limiting values that yield the desired performance. The acceptance limits are the limiting values that permit acceptance of the product. Deciding on the AQL and RQL requires a determination of what is acceptable and unacceptable material.

The same discussion that applied to establishing AQL, RQL, and specification limits for quality characteristics to be used for payment determination applies to characteristics used for screening tests. However, the fact that the sample sizes are likely to be one, or at most two, makes the analyses less involved since PWL, AAD, and CI are not likely to be the quality measure. Because of this, the acceptance limits and specification limits are likely to be the same (another way of looking at this would be to say that there are only specification limits since there is no additional quality measure on which to base acceptance limits) since the measure of quality will usually be the test result or the average of two test results. In theory, an AAD or CI value could be calculated for a single test result so they could be thought of as potential quality measures for screening tests. However, when the sample size is one, calculating AAD or CI is the same process as comparing a single test to a set of specification limits.

With only a single test result on which to decide whether or not to incorporate the material into the project, there is no way to measure variability. Therefore, the screening test is really just a measure of the mean of the quality characteristic for the material being evaluated. The fact that a single test, or the average of two tests, is not a particularly good measure of the mean of a population indicates that screening tests by nature can have potentially high risks. While risks are mentioned briefly in the following examples, risks are addressed in detail in chapter 7.

35.1. Setting Specification Limits for Screening Tests

A simple example will help to illustrate how specification limits might be set for a screening test.

Example

> Suppose that an agency decides to use slump as a screening test for PCC. Guidance for the determination of AQL and RQL for slump might be obtained from agency historical records or from published standards such as those from the ASTM or the American Concrete Institute (ACI).
>
> For example, a slump range of 25 millimeters (mm) to 75 mm might be obtained from *ACI 211*, "Standard Practice for Selecting Proportions for Normal, Heavyweight, and Mass Concrete." Then, specification limits of ±12.5 mm for specified slump of 50 mm or less and ±25 mm for specified slumps of more than 50 mm through 100 mm might be obtained from *ASTM C 94*, "Standard Specification for Ready–Mixed Concrete." A slump of 75 mm might be specified, yielding a lower specification limit of 50 mm and an upper specification limit of 100 mm [i.e., 75 mm ± 25 mm].

Based on the specification limits from the above example, if the result of a screening test for slump fell within the range of 50 mm to 100 mm, the material would be considered acceptable for incorporation into the project. If the slump test were outside of this range, the material would be considered unacceptable for incorporation into the project. Questions regarding whether or not to allow addition of water, or other measures, to bring the concrete mix within the slump requirements would also need to be addressed as part of the technical aspect of the specification.

This shows how engineering decisions can be used to arrive at the specification limits for quality characteristics to be used as screening tests. However, while the process seems simple, the small sample size can lead to high risks of making the wrong decision. Suppose that the records of the agency indicate that 12.5 mm is a typical value for the standard deviation for slump. The following example shows how this information can be used to investigate the risks involved in the screening test.

Example

If the average slump is desired to be 75 mm, and the typical standard deviation for slump is 12.5 mm, i.e., $\sigma = 12.5$, and we assume that slump follows a normal distribution, then the desired (or AQL) population is as indicated in figure 21. The specification limits of 50 mm to 100 mm are also indicated in this figure. It is seen that some of the AQL population falls outside of the specification limits. It is therefore possible that the contractor could produce exactly the desired population, but still have a slump test fall outside of the specification limits, thereby leading to the rejection of the material even though it actually meets the AQL requirement. By calculating the *Z*–values and using the standard normal table in table 8 this possibility can be quantified as follows:

$$Z_{50} = \frac{50-75}{12.5} = -2.0 \quad \text{and} \quad Z_{100} = \frac{100-75}{12.5} = +2.0 \qquad (17)$$

So, the probability of rejecting this population, i.e., the probability of a single test result being either less than 50 mm or greater than 100 mm, is 1.0 minus the area of the normal distribution that is between $Z = -2.0$ and $Z = +2.0$. From table 8, this probability can be calculated as 1.0000 – 0.9544 = 0.0456, or 4.56 percent. This is a risk to the contractor, and is represented by the shaded regions in figure 21.

Now, suppose that the agency decides, based on historical records, engineering calculations, or engineering judgment, that the material should definitely not be accepted for incorporation into the project if the mean slump for the population is 25 mm or more above the target slump of 75 mm. This defines RQL material. An example of an RQL population is shown in figure 22, along with the specification limits of 50 mm to 100 mm. As can be seen from the figure, fully half of the RQL population falls within the specification limits. Therefore, there is a 50 percent probability (this can be verified by calculating the *Z*–value and using table 8) that the RQL population will be accepted by virtue of the screening test for slump yielding a value within the specification limits. This is a risk to the agency, and is represented by the shaded regions in figure 22.

While this is a simple example, it nevertheless clearly illustrates the fact that any single screening test, due to the small sample size, has high risks of accepting rejectable material. It must also be noted that with a single test there is no measure of variability. The single–test screening process therefore is based on the implicit assumption that the standard deviation for the lot is no greater than the one used to derive the specification limits. This may provide a false sense of security that the specification requirements are being met. In reality, at best, screening tests are intended to keep highly nonconforming material from being incorporated into the project. They are neither really intended to nor able to identify material that is “marginally” nonconforming. They will not, therefore, be able to discern RQL material if the definition for RQL is not considerably different from that for AQL material.

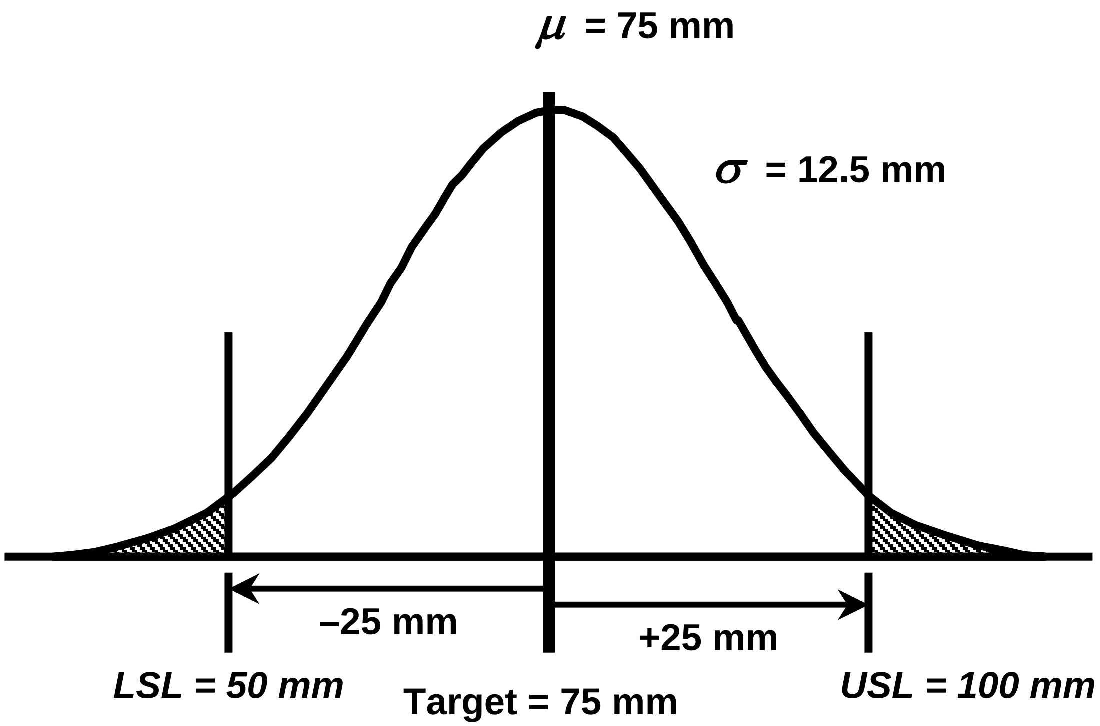


Figure 21. AQL Population for the Screening Test Example

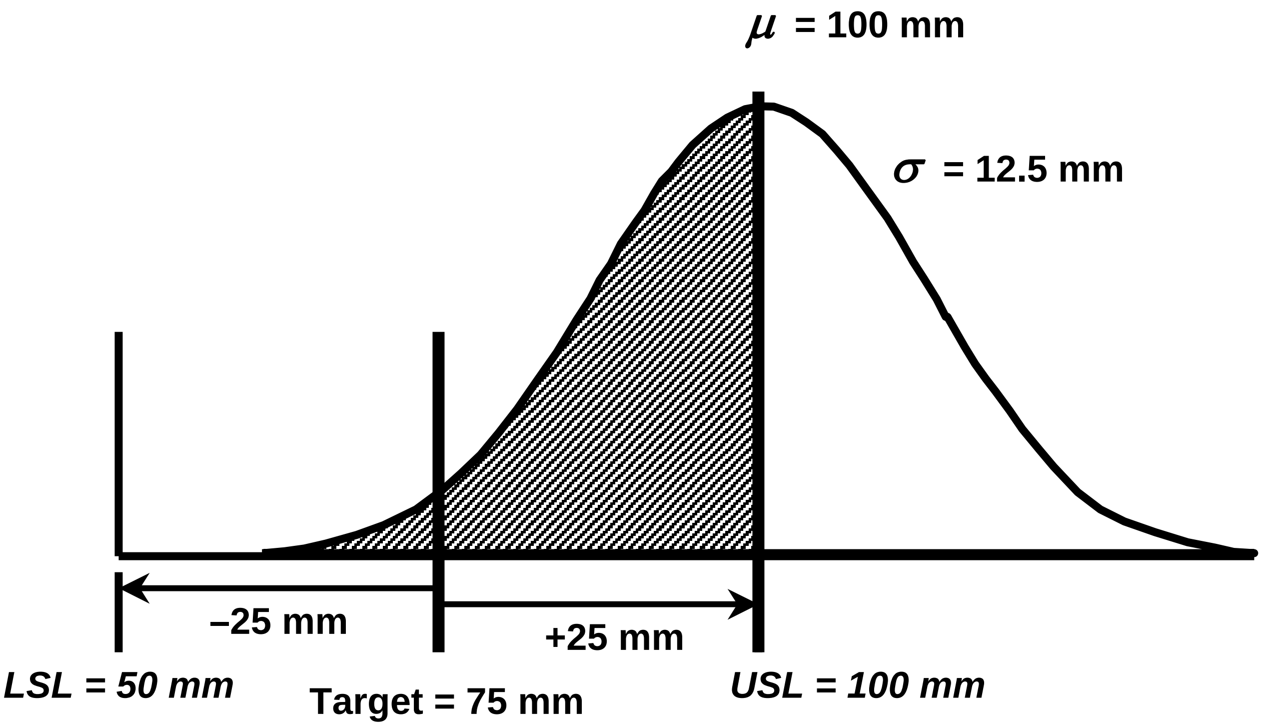


Figure 22. RQL Population for the Screening Test Example

36 Screening Test Quality Characteristics: Determine Acceptance, Rejection, and Rework Procedures

Specific decisions must be made regarding how many test results will be used and how the acceptance/rejection procedures will be used. As has been mentioned several times, the more test results on which to base a decision, the more accurate the decision is likely to be. However, a primary purpose for a screening test is to allow a decision on the quality of the product to be made quickly. Thus, by their nature, screening tests will be few in number. This means, by default, that the risk level may not be as low as desired, but it must certainly still be within the realm of practicality.

By the nature of screening tests, the acceptance and rejection procedures will most likely be to take one test, use the material if the test result is within the allowable specification limits, and do not allow the use of the material if the test result is not within the specification limits. Due to the expense involved if acceptable material is incorrectly rejected, the agency could decide to establish a retest procedure that calls for acceptance if the test result is within certain tolerances, retest if it is outside these tolerances but within a wider set of “reject” tolerances, and rejection if the test result is outside the reject tolerances. If the material is retested without first reworking the material, then it must be stipulated whether the second test will replace or be combined with the first test.

Unless there is a reason to suspect that the first test result is an error, such as the test being run improperly, the first test result should not be discarded and replaced with the retest. This practice has been incorrectly used for many years. However, since the material is tested before it is incorporated into the project, there are occasions in which failing material can be reworked or altered and then retested. If this is feasible, the procedures for reworking and retesting must be detailed. When a product is reworked, it is normally considered to be a new population and, when retested, it is considered as if it had not previously been tested.

37 Payment Quality Characteristics: Decide on the Payment Relationships

There are several decisions that must be made concerning payment relationships. These are extremely important. Experience of the authors has shown payment relationships in an acceptance plan to be the most important factor from a contractor’s perspective. The contractor submits a bid with a certain expectation of the amount of payment for the product. Achieving this amount (or more) of payment is critical in maintaining a viable business. Therefore, maintaining close liaison with the task force that is developing the acceptance plan and keeping the industry apprised of payment factor decisions is imperative when establishing these procedures. Relating quality and performance to payment is the most desirable form of payment relationship because the relationship supports and defends the decision. This is true because negative payment adjustments are typically viewed with skepticism by the contracting industry.

However, when the payment schedule can be shown to be related to quality and, preferably, to performance, it is viewed to be more credible than when it is established arbitrarily.

LCC analyses, which relate quality to performance, are being developed for some materials, and the use of this concept is encouraged. Performance–related payment relationships, such as LCC, require a model relating quality to performance. However, these models may not exist for all properties. Thus, payment relationships other than those that are performance–related are used. These other relationships may be exclusive of or integrated with performance–related payment relationships. These may include the use of incentives/ disincentives, minimum payment provisions, remove and replace provisions, and retest provisions. When used with a payment factor, the AQL should be set such that it yields an expected payment of 100 percent of the unit bid price. When the RQL is used with a payment factor, the agency must decide whether to require removal and replacement or the assignment of a minimum payment factor at the RQL.

The use of incentives for exceptional quality is becoming commonplace practice for many agencies and is viewed as an incentive for the contractor to improve quality. However, the use of incentives is not viewed positively by all agencies. Some think that the use of an incentive is paying extra for what is typical quality. For this reason, it is important to try to assure that the AQL is properly established such that the incentives are applied only for exceptional quality.

37.1. Acceptance by Payment Adjustment—Background

During the latter part of the 20th century the highway profession developed the idea of acceptance of construction work by payment adjustment. This approach foreshadowed the current trend toward performance–related specifications that use mathematical models to predict expected life that, in turn, is combined with LCC analysis to develop appropriate payment schedules.

The strongest argument for this approach is its practicality. While many statistical acceptance procedures used in the private sector tend to characterize a lot as either acceptable or unacceptable, such a sharp distinction is not considered appropriate for most highway construction items. Highway engineers felt more comfortable defining a high level of quality that is clearly acceptable (AQL) and another, substantially lower, level of quality that is clearly rejectable (RQL). In between, the work is not so defective that removal and replacement is required, but neither does it warrant full payment.

Another strong motivating factor was the gradual shift toward end–result specifications. Under the earlier, method–type specifications, agencies had to specify in precise detail how an item was to be constructed, had to devote considerable personnel to inspection activities to make sure the detailed instructions were followed, and, in spite of this, still found that they were often legally responsible if the finished product did not in some way measure up to the desired result. Conversely, with end–result specifications, the agency had the far simpler task of defining a measurable result, and the contractor was given considerable latitude to use its expertise to determine how best to accomplish that result. Besides being a simpler process, this method had the added advantage that it placed the bulk of the responsibility for producing a satisfactory product on the contractor. Since the finished product was evaluated in a quantitative way, this approach lent itself extremely well to the use of adjusted payment schedules to award payment in proportion to the degree to which the desired end result was achieved.

It was eventually realized that, if it made sense to reduce payment for substandard work, it would also make sense to offer some degree of monetary incentive for superior work that exceeds the AQL. Just as the justification for reducing payment for marginally defective work is based on the anticipated increase in future maintenance and repair costs, it was recognized that extra quality usually benefits agencies by reducing these same costs. Therefore, it is justifiable to pass some of these savings back to the conscientious contractor in the form of modest incentive payments in addition to the contract bid price. The incentive payment concept was initially supported by the FHWA as an experimental feature. (19) After several years of satisfactory experience, it was approved for general use and is now a standard feature in many highway construction specifications.

Historically, the payment adjustment approach has been used with a variety of statistical quality measures. In the early 1960's, analysis of data from the AASHO (now AASHTO) Road Test (1958–1960) demonstrated in a dramatic way just how variable most construction characteristics could be. (1) It was soon recognized that construction quality cannot adequately be described by a single point value, but is better characterized as a statistical distribution. It was found that, in a great many cases, the distributions were sufficiently normal that normal curve theory could be used both to describe the quality level desired and to assess the quality level actually achieved. This led many agencies to define the AQL and RQL in terms of PWL, or its counterpart, PD (the percent of the lot falling outside specification limits), both of which are believed to be indicators of performance. A few agencies have used other statistical measures, such as the mean or the AAD, and the CI has also been proposed as a statistical quality measure upon which payment schedules could be based.

Today, nearly all agencies use the payment–adjustment approach for at least some construction items, and an increasing number have begun to use some form of positive–incentive provision.

37.2. Payment Adjustment Rationale

The primary purpose of a payment schedule is to provide sufficient incentive to produce the desired level of quality at the time of initial construction. Effective payment schedules encourage contractors to apply appropriate QC measures to assure that the finished product will equal or exceed the desired level of quality a high percentage of the time. The rationale of the agency is that the small additional cost of good QC practices expended in advance is a better bargain than being faced with the anticipated future costs of poor quality construction, which may lead to premature failure of pavements, excessive maintenance repairs, possibly unsafe driving conditions, etc.

A secondary purpose of the payment schedule is to recoup at least part of the anticipated future costs that are likely to occur when poor quality is received. For a variety of reasons, there will occasionally be times when QC measures are either absent or ineffective, leading to less–than–acceptable work. Provided the work is not too seriously deficient, it usually is both impractical and unnecessary to require removal and replacement, and the better solution in these cases is to accept the work at a reduced price. This is consistent with the legal principle of liquidated damages, a well–established means for recovering losses that are difficult to quantify precisely at the time the contract is executed.

37.2.1. Legal Considerations. In essence, an adjusted payment schedule serves the same purpose as a liquidated damages clause because its function is to state an agreed upon monetary remedy for a breach of contract (i.e., the failure to provide the level of quality specified) for a situation in which the monetary damages are not known precisely and can only be estimated.

It is quite clear that the magnitude of the payment reduction must be reasonably appropriate for the amount of damage actually suffered. This stresses the importance of developing the necessary quality–performance relationships that make it possible to estimate the effects of poor quality. However, this need not be interpreted to mean that the amount of damage must be estimated with great precision. With regard to liquidated damages, the U.S. Supreme Court, in *Wise v. United States*, 249 U.S. 361, 365 (1919), stated:

> *… courts will endeavor, by a construction of the agreement which the parties have made, to ascertain what their intention was when they inserted such a stipulation for payment, of a designated sum or on a designated basis, for a breach of a covenant of their contract. … When that intention is clearly ascertainable from the writing, effect will be given to the provision, as freely as to any other, where damages are uncertain in nature or amount or are difficult of ascertainment or where the amount stipulated for is not so extravagant, or disproportionate to the amount of property loss, as to show that compensation was not the object aimed at or as to imply fraud, mistake, circumvention or oppression. There is no sound reason why persons competent and free to contract may not agree upon this subject as fully as upon any other, or why their agreement, when fairly and understandably entered into with a view to just compensation for the anticipated loss, should not be enforced.*

In simpler language, this states that two contracting parties may agree on the amount to be withheld in the event of noncompliance, and that the courts will uphold this agreement provided that the stipulated amount is reasonably appropriate for the damages actually suffered and there is no element of deception, either consciously or inadvertently. This decision seems to provide solid support for the payment adjustment concept.

Although the liquidated–damages concept has traditionally been applied to losses related to delay of completion, there is no apparent reason why this same rationale should not also apply to losses resulting from a failure to provide the specified level of quality. A logical extension of that argument is that it should also apply to monetary incentives awarded for superior quality. This acknowledgment that extra quality translates into additional value lends credibility to the payment–adjustment concept as a whole.

37.3. Advantages of Positive Incentive (Bonus) Clauses

First, an attempt should be made to demystify the terminology. Particularly in the highway field, in which the idea of receiving extra payment for extra quality is still relatively new, there has been some reluctance to refer to this as a “bonus clause” for fear that this might imply that something is being given away. Those who have had to make a persuasive case for the use of these clauses to top–level administrators, legal counsel, or even legislative bodies, have done so on the basis that these extra payments are not a gift but, in fact, have been earned in return for extra attention paid to QC. The term “bonus” is firmly rooted in the engineering lexicon and, indeed, is frequently heard in the conversations of highway specification writers. It is proposed that either term, bonus or positive incentive, is acceptable, and that the more important consideration is the manner in which such a provision is actually applied.

37.3.1. Fairness Issue. So, whether it is referred to as a positive incentive or a bonus, there are several arguments that can be made in its favor. The first is the matter of fairness. An actual example from a highway agency may help to explain the fairness issue. Several years ago, the agency did not believe that it was necessary to include an incentive clause. This was not based on any type of analysis; it was just a universal opinion that was held at that time.

What changed the agency’s thinking was the field trial of a new specification for PCC compressive strength. The agency had explained to the contractors and suppliers that, to be considered acceptable under this specification, at least 90 percent of the lot must have compressive strengths greater than the class design strength (i.e., the AQL can be expressed as PWL = 90). This was one of the earliest field trials, and payment adjustments were to be computed but not actually assessed. The total project consisted of about $2 million worth of PCC and, when all the results were in, a surprising thing had occurred. The QC had been very consistent, the project was almost exactly at the AQL of 90 PWL, but the average payment factor came out to be 97 percent. In other words, in return for supplying exactly the level of quality that the specification had defined as acceptable, the contractor would have received a payment reduction of about $60,000!

Needless to say, this was brought to the agency’s attention by the construction industry, but that was not necessary. The agency was already scrambling to try to figure out what had gone wrong with an acceptance procedure that was not unlike many others being used around the country at that time. This time an analysis was conducted to see what had happened. The cause of the problem was immediately apparent and was first reported in 1980. (20)

The explanation is quite simple. The standard method for estimating PWL was known to be an unbiased statistical procedure, so that was not the problem. The problem was related to the fact that any statistical estimation procedure has some degree of variability associated with it, particularly at the smaller sample sizes conventionally used for highway construction specifications. In other words, while the average of a large quantity of estimates will be very close to the true population mean value (which happened to be the AQL of PWL = 90 in this particular case), the individual estimates will be both above and below it to varying degrees. Those that were below PWL = 90 all received some degree of payment reduction, but all those above 90 PWL were limited to the maximum payment factor of 100 percent. Obviously, this will cause the average payment factor to be biased downward and, in this particular field trial, it produced an unwarranted payment reduction of about $60,000.

Just as the explanation is simple, so is the solution. Even a relatively small positive incentive (or bonus) for work exceeding the specified AQL will correct this situation. In all cases, the OC curve should be constructed to confirm that the acceptance procedure is working as desired and, in particular, that the average payment factor at the AQL is 100 percent. The subject of OC curves is addressed in detail in chapter 7.

37.3.2. Effect on Quality. The introduction of a new acceptance procedure with a payment adjustment clause fully in effect tends to produce significant increases in the level of quality received. While there have been no controlled studies to attempt to separate the effect of a bonus provision from the effect of the payment schedule as a whole, it seems quite likely that a bonus provision can only serve to enhance the motivation to produce good work.

37.3.3. Construction Industry Relations. Virtually every agency that has introduced statistical acceptance procedures with adjusted payment schedules has had to overcome considerable resistance from the construction industry. Although much of the resistance can be attributed to a general fear of the unknown, there is no denying that specifications of this type can impact severely on a contractor’s means of livelihood. It is not difficult to understand why there might be strong resistance to a system that is perceived to be only punitive, that only penalizes for poor performance but does not reward excellent performance. The most effective way to counter this perception is to make it possible to earn tangible rewards for superior performance. The inclusion of positive incentives (or bonuses) casts the whole system in a new light and is especially appealing to those contractors who have made QC an integral part of their operations.

37.3.4. Economic Benefits. From an economic standpoint, a bonus provision can be claimed to be beneficial if the added value of increased quality produced by the bonus provision exceeds the amount of money paid out in actual bonuses. More precisely, the added value only has to exceed the amount by which the total contract cost was increased.

Although it is nearly impossible to find suitable projects to serve as test cases and controls to quantify the effect of bonus provisions, it is easy to show (see appendix I) that only modest increases in expected pavement life would be necessary to justify the use of bonus clauses.

However, a word of caution is in order regarding the magnitude of bonus payments. Because there may be other modes of failure besides those involving the quality characteristic for which the bonus is being paid, many agencies have either identified a maximum amount of bonus, or else have made the bonus payment conditional upon satisfactory completion of other parts of the contract, or both.

37.3.5. Effect on Bidding Process. There appears to be growing evidence that there is yet another benefit of bonus provisions that was not initially anticipated or recognized. A construction firm that has made QC an integral part of its operation may be able to submit a lower bid because it is confident that it can earn a substantial bonus. If this effect is found to be typical, the use of bonus provisions will be an effective way for agencies, which are bound by the competitive bidding system, to put more work in the hands of highly qualified contractors.

37.4. Types of Payment Schedules

The earliest payment schedules were usually stepped schedules, such as that shown in table 12 and plotted in figure 23.

Table 12. Typical Stepped Payment Schedule Based on PWL

Estimated PWL	Payment Factor, %
95.0 — 100.0	102
85.0 — 94.9	100
50.0 — 84.9	90
0.0 — 49.9	70

More recently, there has been a tendency to use continuous (equation–type) payment schedules such as that shown in the equation 18 below and also plotted in figure 20.

$$PF = 55 + 0.5\ \mathrm{PWL} \tag{18}$$

Where: PF = payment factor as a percent of contract price.
PWL = estimated percent within limits.

Although risk analysis (see chapter 7) would show these two payment schedules to have very nearly the same long–term performance, there is a distinct advantage associated with the continuous form. When the true quality level of the work happens to lie close to a boundary in a stepped payment schedule, the quality estimate obtained from the sample may fall on either side of the boundary due primarily to chance. Depending upon which side of the boundary the estimate falls, there may be a substantial difference in payment level, which may lead to disputes over measurement precision, round–off rules, and so forth. This potential problem can be completely avoided with continuous payment schedules that provide a smooth progression of payment as the quality measure varies.

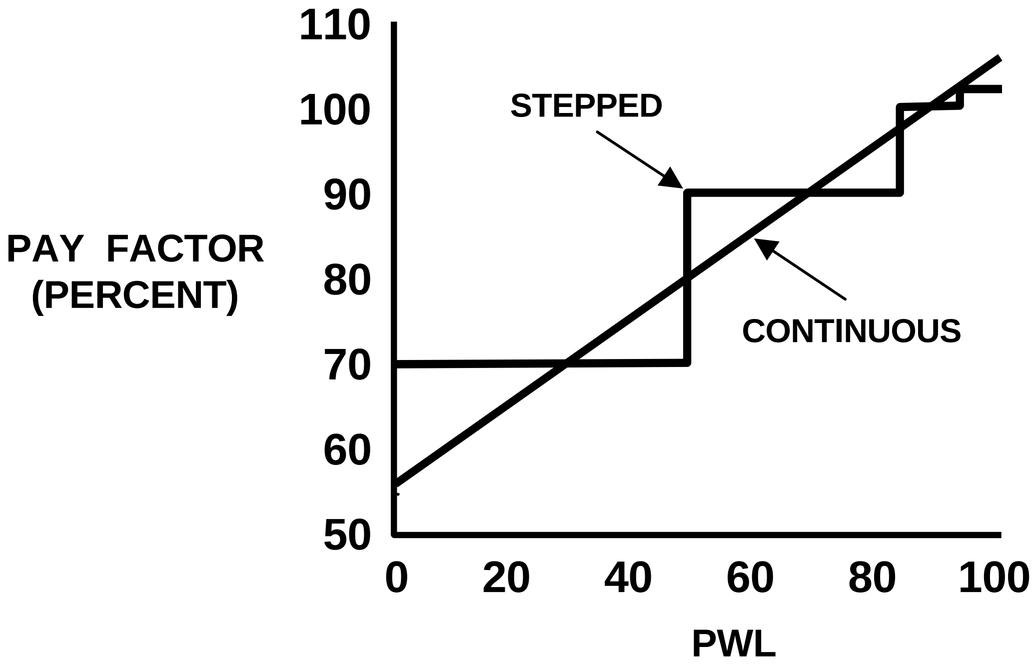


Figure 23. Example of Stepped and Continuous Payment Schedules

37.5. Life–Cycle Cost (LCC) Basis For Payment Adjustments

Ordinarily, a pavement is designed to sustain a specified number of load applications before major repair (such as resurfacing) is required. If, due to construction deficiencies, the pavement is not capable of withstanding the design loading, it will fail prematurely. The necessity of repairing this pavement at an earlier date results in an additional expense that, since it usually occurs long after any contractual obligations have expired, must be borne by the agency. Therefore, one possible purpose for an adjusted payment schedule might be to withhold sufficient payment at the time of construction to cover the extra cost anticipated in the future as the result of deficient quality work.

Pavements are usually designed to withstand a required number of equivalent single axle loads (ESALs). For those quality characteristics used in the design procedure, the as–built values can be compared to the design values to estimate the fraction of design loadings the pavement is capable of sustaining. As an approximate estimate, this fraction can be multiplied by the design life to obtain the expected life of the pavement. If greater precision is desired, a traffic growth rate can be assumed, the effect of which is to extend slightly the expected life (since fewer of the allowable loads will occur in the early part of a pavement’s life).

To estimate the cost to the agency of premature pavement failure, it is necessary to determine the net present value of the various actions made necessary by early failure. Unlike various intuitive methods that often require unrealistic assumptions, the LCC basis presented in this section assumes a practical repair strategy that an agency may employ.

For example, suppose that experience has shown that overlays typically last about 10 years. If the initial resurfacing were to fail one or two years prematurely, it is not likely that an agency would do a minor repair to extend the life of the pavement to the originally expected value of 10 years. A much more practical decision would be to reschedule the overlay that was planned for the 10th year and do it one or two years sooner. However, if the 10th year overlay is rescheduled to an earlier date, and overlays typically last 10 years, then all future overlays must as well be moved earlier in time.

Because it provides a valid and rational way to estimate the net present value of future actions such as these, LCC analysis makes it possible to obtain a realistic estimate of the cost of the actions resulting from premature failure. The procedure involves the calculation of a series of debits and credits and turns out to be relatively easy. Moving the 10th year overlay to the 8th year, for example, would result in a debit in net present value terms because it represents a cost in the 8th year that was not planned. However, there will also be a credit for no longer having to do an overlay in the 10th year. Since the 10th year overlay is farther in the future, the credit for this action is discounted to a greater degree, resulting in a net debit for the rescheduling of the 10th year overlay. While it is true that the net debits from the rescheduling of overlays farther in the future are discounted to a greater extent, and soon become insignificant, ignoring them altogether would substantially underestimate the true cost of pavement failure. Alternatively, selecting a specific analysis period would require an assumption about the residual value of a partially depleted overlay, information that is not readily available.

Fortunately, this is an easy problem to solve mathematically and the derivation is given in appendix I. This produces the expression given in equation 19, which requires input information that is readily available or can easily be obtained.

$$PAYADJ = \frac{C(R^D - R^E)}{(1 - R^O)} \qquad (19)$$

where: $PAYADJ$ = appropriate payment adjustment for new pavement or overlay (same units as C).

C = present total cost of resurfacing. (typical value = \$23.92/m^2 (\$20/sy)).

D = design life of pavement or initial overlay. (typically 20 years for new pavement, 10 years for overlay).

E = expected life of pavement or overlay (independent variable).

O = expected life of successive overlays (typically 10 years).

R = $(1 + INF) / (1 + INT)$.

INF = long–term annual inflation rate in decimal form (typically 0.04).

INT = long–term annual interest rate in decimal form (typically 0.08).

Example

Suppose that, based on an appropriate performance relationship, the as–constructed resurfacing, for which an appropriate payment adjustment is to be determined, is expected to last *E* = 8 years instead of the design value of *D* = 10 years. The cost of this premature failure would be estimated by first computing *R* = 1.04 / 1.08 = 0.963 and then applying equation 18, as follows:

$$\boldsymbol{PAYADJ} = \frac{\$23.92\mathbf{(}0.963^{10} - 0.963^{8}\mathbf{)}}{\mathbf{(}1 - 0.963^{10}\mathbf{)}} = \mathbf{–\$}4.09\,/\,\mathbf{m}^{2}$$

In the same way, the appropriate payment adjustments for other values of expected life can be calculated as summarized in table 13.

Table 13. Calculation of *PAYADJ* for Selected Levels of Expected Life

Expected Life, years	**Appropriate Payment Adjustment**
	$/m^2
12	+3.79
10 (Design)	0.00
8	–4.09
6	–8.50
4	–13.27
2	–18.40
0	–23.92

It is seen from table 13 that the cost of premature failure can be substantial, terminating at the initial cost of resurfacing of $23.92 per square meter (m^2) for zero expected life. It is common practice for most agencies to make use of an RQL provision that gives them the option to require removal and replacement at no additional expense when the quality falls below some predetermined, seriously deficient level. Such a provision would probably apply before the lower portion of this table is reached, but if for some reason an agency elected not to require removal and replacement, this method provides the levels of payment adjustment that would be justified for extremely poor quality.

It can also be seen from this table that the method properly recognizes that a tangible benefit results when the as–constructed quality exceeds the design standard and extends the expected life of the pavement, thus justifying incentive payments for superior quality.

37.6. Performance Relationships

To apply the LCC basis for payment schedules in a manner that is both fair to all parties and legally defensible, it is necessary to have at least an approximate performance relationship. The purpose of the performance relationship is to predict from quality characteristics measured at the jobsite what the expected service life of the construction item will be. This is the independent variable to be entered into the LCC equation presented in the previous section.

Research is currently in progress to develop mathematical models upon which valid performance–related specifications can be based, and some prototype models are in the development stage. However, it may still require several years before these models are available for widespread implementation. This does not mean that the development of useful and effective payment schedules must be put on hold until these models become available. In many cases, current knowledge, combined with engineering and mathematical principles, may be sufficient to develop interim performance models that can be demonstrated to be serviceable.

37.6.1. Polynomial Model. Prototype models can be developed by identifying a sufficient number of "known" points that are then used to develop generalized mathematical models. Appropriate assumptions must be made concerning mathematical form, and engineering considerations are used to establish realistic boundary conditions. The procedure is presented in sufficient detail in appendix J to allow the reader the opportunity to use different known values or assumptions to develop models that are suited for specific applications.

For example, table 14 illustrates a typical performance matrix that might be used to develop the following type of polynomial performance model for two quality characteristics.

$$EXPLIF = C_0 + C_1 \times \mathrm{PD}_{VOIDS} + C_2 \times \mathrm{PD}_{THICK} + C_3 \times \mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK} \qquad (20)$$

where: $EXPLIF$ = expected life (years).
PD_{VOIDS} = air voids percent defective.
PD_{THICK} = thickness percent defective.
C_i terms = coefficients to be determined.

Table 14. Typical Matrix of Expected Pavement Lives Used to Develop Polynomial Performance Model

	Thickness Quality	
Air Voids Quality	PD = 10	PD = 90
PD = 10	20 yrs.	10 yrs.
PD = 75	10 yrs.	5 yrs.

The method involves using the four pieces of data in table 14 to write four simultaneous equations that can be solved to obtain the four unknown equation coefficients, producing the performance model given in equation 21 below. (The complete development is contained in appendix J.) Values of *EXPLIF* for selected combinations of PD_{VOIDS} and PD_{THICK} have been calculated and are presented in table 15.

$$EXPLIF = 22.9 - 0.163\, PD_{VOIDS} - 0.135\, PD_{THICK} + 0.000961\, PD_{VOIDS} \times PD_{THICK} \qquad (21)$$

Table 15. Values Calculated for Expected Pavement Life

PD_{VOIDS}	PD_{THICK}	***EXPLIF*, yrs.**
0	0	22.9
10 (AQL)	10 (AQL)	20.0 (Design)
100	100	2.7

It can be seen from table 15 that when both quality measures are at the AQL value of PD = 10, the expected life equals the design life of 20 years. For excellent quality with $PD_{VOIDS} = PD_{THICK} = 0$, the *EXPLIF* equation predicts that the pavement life will be extended to almost 23 years, while, for extremely poor quality with $PD_{VOIDS} = PD_{THICK} = 100$, the expected life is reduced to less than 3 years. For any combination of PD_{VOIDS} and PD_{THICK} in between, the *EXPLIF* equation gives an appropriate estimate for expected life.

The accuracy of this equation is obviously dependent upon the use of realistic values for expected life in the performance matrix in table 14. The values used in the example in this table have been estimated by the New Jersey Department of Transportation (NJDOT), which believes that the resultant values for expected life predicted by the performance model are realistic. This model has been used as the basis for the composite quality measure (described in appendix K) used by the NJDOT in their current acceptance procedures for HMAC pavement (see case study 2 in chapter 9).

37.6.2. Exponential Model. Although the above polynomial procedure is easy to apply and can produce a very serviceable model when only two quality characteristics are involved, another approach is more effective when two or more quality measures must be included in the acceptance procedure. A detailed presentation of this method is contained in appendix L and involves the use of the exponential model in equation 22.

$$\boldsymbol{EXPLIF} = \boldsymbol{A}\boldsymbol{e}^{-(\boldsymbol{B}_1 \mathrm{PD}_1 + \boldsymbol{B}_2 \mathrm{PD}_2 + \cdots + \boldsymbol{B}_{\boldsymbol{k}} \mathrm{PD}_{\boldsymbol{k}})} \qquad (22)$$

where:
EXPLIF = expected life, years.
A = constant to be determined.
B_k = coefficients to be determined for each of the *k* quality characteristics.
PD_i = percent defective of individual quality characteristics.
k = number of quality characteristics.
e = base of natural logarithms.

It is explained in Appendix L why PD is better suited than PWL as the statistical quality measure for this particular model. However, PWL can be used with this model simply by substituting (100 – PWL) for each of the PD terms, if desired.

This model has certain important advantages. It tends to produce a sigmoidal ("S") shape that is believed to be an appropriate form for many performance relationships. Also, because this particular model form produces a maximum of "*A*" and a minimum as close to zero as desired (but not below zero), it can easily be made to fit most real–world situations. Finally, it requires relatively straightforward data and simple mathematics to accommodate as many acceptance characteristics as are likely to be necessary. (A detailed example involving in–place air voids, thickness, and smoothness of HMAC pavement is presented in appendix L.)

If the method is to be valid, it must be based on realistic data, and if it is to be practical, the required data must be readily obtainable. Table 16 is a generic data matrix that must be completed to develop the exponential model.

Table 16. Data Matrix to Develop Exponential Performance Model

$\mathbf{PD}_1$	$\mathbf{PD}_2$	$\mathbf{PD}_3$	→	$\mathbf{PD}_k$	***EXPLIF***
AQL(1)	AQL(2)	AQL(3)	→	AQL(*k*)	*DESLIF*
POOR(1)	AQL(2)	AQL(3)	→	AQL(*k*)	*LIFE*(1)
AQL(1)	POOR(2)	AQL(3)	→	AQL(*k*)	*LIFE*(2)
AQL(1)	AQL(2)	POOR(3)	→	AQL(*k*)	*LIFE*(3)
↓	↓	↓	→	↓	↓
AQL(1)	AQL(2)	AQL(3)	→	POOR(*k*)	*LIFE*(*n*)

PD_i = percent defective for each of the *k* quality characteristics.
EXPLIF = expected life in years.
DESLIF = design life in years.
AQL(*i*) = acceptable quality level, in PD, for each of the *k* quality characteristics.
POOR(*i*) = poor quality level, in PD, for each of the *k* quality characteristics.
LIFE(*m*) = expected life in years for *n* selected combinations of PD levels.

For example, consider a resurfacing project for which historical data have shown the typical expected life to be about 10 years. A typical value for the AQL is PD = 10, while RQL values tend to vary more widely, depending on what quality level the agency believes justifies removal and replacement at the contractor's expense. For purposes of this example, suppose the agency has decided to use RQL values of PD = 65, 75, and 85, respectively, because it is believed that these levels correspond to approximately a 50 percent loss of pavement life, or an expected life of 5 years. These assumptions lead to the completed data matrix shown in table 17.

Table 17. Completed Data Matrix for Example of Exponential Model

PD_{VOIDS}	PD_{THICK}	PD_{SMOOTH}	***EXPLIF*, in years**
10 (AQL)	10 (AQL)	10 (AQL)	10 (AQL)
65 (RQL)	10 (AQL)	10 (AQL)	5 (poor voids)
10 (AQL)	75 (RQL)	10 (AQL)	5 (poor thickness)
10 (AQL)	10 (AQL)	85 (RQL)	5 (poor smoothness)

The ease of applying this method should now be apparent. All that remains is to use the information in the data matrix to solve for the unknown coefficients in the exponential performance equation for *EXPLIF*. To accomplish this, it is first necessary to take logarithms of both sides, producing equation 23:

$$\ln(EXPLIF) = \ln(A) - B_1\mathrm{PD}_{VOIDS} - B_2\mathrm{PD}_{THICK} - B_3\mathrm{PD}_{SMOOTH} \qquad (23)$$

where:
$EXPLIF$ = expected life, in years.
PD_{VOIDS} = air voids percent defective.
PD_{THICK} = thickness percent defective.
PD_{SMOOTH} = smoothness percent defective.
A, B_1, B_2, B_3 = unknown coefficients.
ln = natural logarithm operator.

The complete set of equations is presented in Appendix L along with the solution, leading to the following performance model:

$$EXPLIF = 13.8e^{-(0.0126\mathrm{PD}_{VOIDS} + 0.0107\mathrm{PD}_{THICK} + 0.00924\mathrm{PD}_{SMOOTH})} \qquad (24)$$

A lengthy series of tests is described in Appendix L to confirm that the model will produce realistic values of expected life for any combination of quality levels in the three measures. Once these checks have been completed, the model can be relied upon to produce the value of *EXPLIF* needed for the LCC equation, which is equation 19.

37.7. Composite Payment Factors

Most specifications will contain multiple quality characteristics. How can these be combined to come up with a single payment factor for a lot? The ideal situation is to have a performance model that can predict long–term performance of the payment. In such a case, the quality characteristic measurements can be input to the model and the payment factor can be based on the predicted performance of the in–place pavement as compared to the desired performance. Unfortunately, such models are either not available or are not widely accepted at this time. Therefore, other methods for determining the composite payment factor are currently in use.

Various agencies have considered at least four different approaches for combining a number of payment factors for individual acceptance quality characteristics into a single composite payment factor. These approaches include:

- Using the minimum individual payment factor.
- Averaging (possibly with weighting factors) the individual payment factors.
- Multiplying the individual payment factors.
- Summing the individual payment adjustments.

The approach using the minimum individual payment factor for the composite is based on the "weak link" theory, i.e., the lowest payment factor indicates the value of all the quality characteristics. For the other three approaches the concept is that all individual factors contribute to the total. However, the composite payment for the three approaches can be quite different depending on the value of the individual payment factors.

An example will help to show how composite payment factors can be determined from individual payment factors. Suppose that for a PCC pavement, the quality characteristics are compressive strength, permeability, and thickness. The composite payment factor determined by various methods can be quite different depending on the magnitude of each individual payment factor. Table 18 shows the composite payment factors for various methods for combining the individual payment factors.

Table 18. Examples of Methods for Combining Individual Payment Factors Into a Composite Payment Factor

Individual Payment Factor			Composite Payment Factor			
Strength	**Permeability**	**Thickness**	**Minimum**	**Average**	**Multiply**	**Sum**
1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.05	1.05	1.05	1.05	1.05	1.16	1.15
0.80	0.80	0.80	0.80	0.80	0.51	0.40
1.00	0.80	1.05	0.80	0.95	0.84	0.85

The averaging, multiplying, and summing methods for combining individual payment factors implicitly assume that each individual payment property is equally important. Several agencies have chosen to weight the payment factors with the concept that some quality characteristics are more important than others. For HMAC, when mixture properties and field compaction are used as quality characteristics, the in–place air voids are often weighted more heavily than the mixture properties.

For example, a weighting system that is used by one agency is as follows:

$$PF = 0.2AV + 0.1VMA + 0.1AC + 0.6DEN \tag{25}$$

where: PF = composite payment factor.
AV = payment factor for laboratory–compacted air voids.
VMA = payment factor for voids in mineral aggregates.
AC = payment factor for asphalt content.
DEN = payment factor for field in–place density.

Another agency uses a different weighting system for the same HMAC properties:

$$PF = 0.35AV + 0.1VMA + 0.2AC + 0.35DEN \tag{26}$$

Weighted average composite payment factors such as these are intuitively appealing since it is very likely that all payment quality characteristics do not have the identical impact on pavement performance. A drawback to this approach, however, is that there is no obvious methodology for determining the appropriate weightings. The weightings, therefore, are subjective in nature and, as the above equations show, will hence vary from agency–to–agency depending upon agency or individual preferences.

37.8. Composite Quality Measures

As noted in the previous section, statistical construction specifications based on multiple quality characteristics frequently use payment equations that include a separate term for each of the quality characteristics so that the resultant payment adjustment is a function of the combined effect of all quality measures. An alternate method to accomplish the same purpose is to base the payment equation on a single quality measure that is a composite of the individual quality measures. This latter approach, because it keys the various decisionmaking steps to a single performance indicator, simplifies the procedure and offers several practical advantages.

For example, the performance relationship that was presented previously for expected pavement life can be used to develop a single, composite quality measure upon which all of the various acceptance decisions (accept, reject, retest, payment adjustment, etc.) can be based. Equation 21 for expected pavement life, which was developed in a previous section above, is as follows:

$$EXPLIF = 22.9 - 0.163\ \mathrm{PD}_{VOIDS} - 0.135\ \mathrm{PD}_{THICK} + 0.000961\ \mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK} \tag{21}$$

As described in detail in appendix K, a simple transformation converts this equation into the composite quality measure given by equation 27:

$$\mathrm{PD}^* = 0.807\ \mathrm{PD}_{VOIDS} + 0.669\ \mathrm{PD}_{THICK} - 0.00476\ \mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK} \tag{27}$$

where: PD^* = composite quality measure in units of percent defective.
PD_{VOIDS} = air voids percent defective.
PD_{THICK} = thickness percent defective.

PD* progresses smoothly from zero to 100 percent as the individual quality measures, PD_{VOIDS} and PD_{THICK}, vary throughout the same range. Table 19 presents a few selected examples of this. More extensive tables and graphs are contained in appendix K.

Table 19. Examples of Computed PD* Values for Selected Individual PD Values

PD_{VOIDS}	PD_{THICK}	***EXPLIF*, yrs.**	***PD****
0	0	22.9	0.0
10	10	20.0	14.3
50	50	10.4	61.9
25	75	10.5	61.4
100	100	2.7	100.0

It can be seen in table 19 that the case in which PD_{VOIDS} and PD_{THICK} are both equal to 50 produces essentially the same level of expected life as the case in which $PD_{VOIDS} = 25$ and $PD_{THICK} = 75$. This result flows directly from the manner in which the *EXPLIF* equation was derived and is realistic because an increase in the quality of one measure might be expected to offset a decrease of quality in the other. Appropriately, both cases produce virtually the same value of PD* in the last column of the table, indicating that PD* is well–suited as a measure upon which a QA specification can be based.

This property of the composite quality measure, which properly accounts for the combined effect of multiple quality characteristics, also makes it possible to develop an RQL provision that is far superior to the alternative of defining separate RQL provisions for the individual quality measures. For the example in table 20 it is assumed that the agency has defined for air voids and thickness separate RQL provisions of $PD_{VOIDS} \geq 75$ and $PD_{THICK} \geq 90$. Clearly, case 3 in table 20 is by far the worst case, yet it is not recognized as an RQL condition when using individual RQL provisions, while the other two cases are.

Table 20. Illustration of Problem with Separate RQL Provisions

Case	**Quality Level**		**Reject?**	**PD***
	Air Voids	**Thickness**		
1	PD = 75 (RQL)	PD = 0 (Excellent)	Yes	60.5
2	PD = 0 (Excellent)	PD = 90 (RQL)	Yes	60.2
3	PD = 74 (Almost RQL)	PD = 89 (Almost RQL)	No	87.9

To demonstrate the effectiveness of an RQL provision based on the composite quality measure, equation 27 was used to compute the corresponding values for PD* that appear in the last column of table 20. In this example, a PD* value of 60, or more, would be regarded as rejectable and, as can be seen in the last column, case 3 is properly recognized as being well into the rejectable region. The actual development of RQL provisions based on multiple quality measures is covered in greater detail in appendix K.

37.9. Retesting Provisions

There are several reasons why an agency might choose to perform a second set of tests before making the final decision concerning the acceptability or rejectability of a construction item:

- To confirm that the work is truly defective before imposing a severe consequence, such as requiring removal and replacement, or assigning a minimum payment factor.
- To produce a more desirable OC curve that properly balances the risks between the agency and the contractor.
- To guard against a possible breakdown in any of the steps of the sampling and testing process.
- To make more efficient use of limited sampling and testing resources. A reduced sampling effort may be sufficient to identify work that is clearly acceptable or clearly rejectable. Only when the work falls in between these two extremes is the added discriminating power of an increased sample size required.

If a retest provision is to be used, it must be spelled out clearly in the contract documents, including precisely how the retest results are to be processed. There are two distinctly different ways to do this:

- They may be combined with the original test results.
- They may be used in place of the original test results.

Advocates of the first method argue that it makes maximum use of the available information and that it is wasteful to discard any valid information. Advocates of the second method would question whether the original sample was truly valid. If the low quality level were the result of some malfunction of the testing process, then it would be more appropriate to disregard the questionable data.

Each agency must decide for itself which method is more appropriate in any particular situation, depending on the quality characteristic that is being measured and the test method that is being used. If the decision is made to combine the retest results with those obtained from the initial sample, caution must be exercised in computing the OC curve for this procedure since the probabilities of failing the original test and passing the retest are not statistically independent, but are correlated to some degree.

37.10. Double, Multiple or Sequential Sampling

Most construction and materials acceptance plans are what can be called “single sampling” plans. That is, a single sample of a stipulated size is taken and a decision is made based on the test results from this sample. The preceding discussion on retesting indicated some reasons that additional sampling and testing might be conducted. Sampling plans based on more than one level of sampling and testing have been used in manufacturing applications for many years. They are typically based on “accept or reject” attributes acceptance plans rather than the payment adjustment variables acceptance plans that are predominate in highway construction.

37.10.1. What Are They? In “double sampling” plans it is possible that a decision could be made after a smaller first sample is taken and tested, but the decision may be deferred until a second sample is obtained and tested. For example, such a plan might be phrased as follows:

> Take a sample of 5 items and test them for the desired quality characteristic. Accept the lot if there are 0 defective items in the sample. Reject the lot if there are 3 or more defective items in the sample. If there are 1 or 2 defective items in the sample, take an additional sample of 5 new items. Accept the lot if there are 3 or fewer defective items among the cumulative sample of 10 items. Reject the lot if there are 4 or more defective items among the cumulative sample of 10 items.

To further reduce the amount of sampling and testing done, it is customary to curtail testing when the rejection number of defects is attained. For the above acceptance plan for example, if the first sample had 2 defective items then a second sample would be obtained. If 2 additional defective items were obtained after testing 3 items from the second sample, then testing would cease and the lot would be rejected because the reject number of 4 had been reached.

The concept of double sampling can be extended to “multiple sampling,” which represents the case when three or more samples might be obtained before a final decision on the lot is obtained. Multiple sampling acceptance plans may allow for as many as seven samples before a final decision must be made.

“Sequential sampling” is generally used to represent the case where a decision is possible after each item is tested and there are no specified limits on the total number of items that will be tested. Sequential sampling acceptance plans have been used in manufacturing operations primarily when a destructive testing method is required. The underlying purpose behind sequential sampling acceptance plans is to keep to a minimum the number of items that must be destructively tested and still be consistent with the risk levels that are desired. Their use in highway construction may be questionable given the amount of time that it often takes to obtain a test result once a sample has been obtained.

37.10.2. Applications in Highway Construction. While various double and multiple sampling plans have been developed for manufacturing operations, there are some limitations to their use for highway construction and materials acceptance plans. In manufacturing operations the product that is being produced is generally easy to sample and test, and usually does not change properties from the first sample to the second sample taken from the lot. For example, if a sample of 10 bolts is obtained and tested from a lot of 300 bolts, there is no reason to believe that the lot will have changed in any way prior to obtaining the second required sample of 10 bolts. This may not be true for many highway construction materials.

The use of double or multiple sampling may be limited in highway construction due to the nature of the way in which samples are obtained and tested. For highway construction it is quite possible, even likely, that the population to be sampled could change between the first and second sample. These potential changes in the population must also be considered if a retest provision is selected for use in the acceptance plan.

For example, if HMAC samples for asphalt content determination are obtained from the back of the truck, production for the lot may be completed by the time the test results from the first sample are available. It would therefore not be possible to obtain a second, or “double,” sample from trucks on this lot. It would also probably not be appropriate to take cores from the pavement to represent the second sample since there is no reason to believe that cored samples will have the same sampling and testing variabilities as sampling and testing from the truck. Remember that chapter 5 stressed the importance of developing specification limits based on the appropriate process variability. Specification limits based on sampling and testing from the truck would not be appropriate for use with sampling and testing of cores.

Similarly, if acceptance were based on 28–day compressive strengths of cylinders cast when the concrete was placed, the only way to obtain a “double” sample would be to cast all of the cylinders for both samples when the concrete was initially placed. It is questionable whether a “second” sample could be obtained by coring and testing the cores since these are such different processes.

There may be cases where it would be possible to develop double sampling acceptance plans, but these must be carefully scrutinized to ensure that the two samples are indeed from the same population. For example, suppose that an HMAC overlay is cored for in–place air voids or density determination, and is then opened to traffic. If the first sample results were not in the “accept” range, would a second set of cores obtained a week later still represent the same population, or would compaction under traffic loading make this a different population? It is likely that the specification limits for this characteristic would have been developed based on the results from cores that had not been subjected to traffic loading.

The double, multiple, and sequential acceptance plans that have been used in manufacturing have been primarily, if not exclusively, for accept or reject decisions. They have not been applied in a situation, such as highway construction and materials, where there may also be an option to accept the lot at an adjusted payment. While the risks and expected payments can conceivably be developed for these more complex procedures, it would be quite a bit more complicated than the case for single sampling plans.

The complications with using more than a single sample that are discussed in this section may be among the reasons that agencies have almost exclusively used single sample acceptance plans for highway construction and materials. While double, multiple, or sequential sampling acceptance plans could possibly be considered for highway construction materials, the procedures must be developed with caution, and some serious degree of reservation due to the factors discussed in the preceding paragraphs.

37.11. Alternatives to Payment Adjustments—Correct or Repair

The typical approach to construction or material that is deficient in quality has been to require it to be removed and replaced if the quality is too deficient, or to accept it at a reduced payment if the deficiency is such that it would not be economically justifiable to require removal and replacement. In some instances, correction or repair of the deficiency might be an additional option.

For example, if coring indicates a deficiency in thickness for an HMAC overlay, then it may be possible to increase the thickness with an additional overlay. Similarly, a pavement with sections that are deficient in smoothness may be corrected by selective grinding. If correction and repair are to be considered as options, then the agency needs to address how to incorporate this into the acceptance procedures, and whether or not it will have any impact on the way in which samples are obtained and their results are analyzed.

For example, one approach that some agencies have used with respect to pavement thickness is to further investigate any coring location for which the thickness is deficient to a certain extent or greater. In the event of a core that is deficient in thickness, some specifications call for additional coring, moving outward from the original core location, to determine the extent of the deficient thickness. The deficient section can then either be corrected or removed and replaced. Similarly, the extent of smoothness deficiencies could be required to be identified and corrected.

If a correct or repair strategy is employed, the agency must be careful regarding how this information is treated with respect to other test results. For example, in the event that a deficient core thickness is identified and then removed and replaced, that core should not be included with other cores from the sample when calculating averages, standard deviations, or PWL or PD values. In fact, if the intent of the coring is to identify areas of deficient thickness, then a random sampling procedure to estimate population parameters may not be the most appropriate sampling approach. A systematic sampling scheme, such as taking a core every 300 meters or 1,000 feet, might be a better way to look for regions that are deficient in thickness. If this is done, however, it must be realized that the data obtained in this systematic manner would not be appropriate to use for estimating population parameters such as mean and standard deviation.

38

Determine Appropriate Lot Size, Sublot Size, and Sample Size

A lot is the amount of material that is to be judged acceptable or unacceptable on the basis of a sample comprised of a stated number of test results. The determination of lot size is primarily an economic decision. Since each lot is tested, for very small lots the cost of testing may exceed the benefits. Very large lots may allow acceptance of a large amount of less than desirable quality or severely negatively impact the price received by the contractor. However, with the increase in the contractor doing both QC and acceptance testing, the likelihood of this occurring is diminished. This is because the contractor most likely will monitor the payment factor concurrently with the quality measure to assure that less than desirable quality, and consequently low payment factors, does not occur. The consequence of this is that some agencies are starting to use relatively large lot sizes. For example, the California DOT and the FHWA Federal Lands Highway Division each use a project or an entire item of construction as a lot.

38.1. Defining a Lot

Lots can be established based on time, on quantity (e.g., tonnage, area, or volume), or on an entire construction/material item for the total project. Each choice has advantages and disadvantages. It is important for the agency to recognize the predominant contractor operation and use that knowledge to select the best definition for a lot.

As an example, a day’s production is one choice for a lot size. The advantage is that the operation goes through a complete cycle of start–up, run, and shut down. The disadvantage is that the quantity of material included in each lot may vary considerably from lot–to–lot because of production interruptions caused by inclement weather, equipment breakdowns, etc.

Another example for a lot size might be a specified tonnage such as 1800 megagrams (Mg). This has the advantage of a consistent amount of material in each lot, while the disadvantage is that each lot may have a different number of production start–up, run, and shut down cycles. Ideally, a lot should represent a single population. That is, the material in the lot should have all been produced from essentially the same process under essentially the same production conditions. This is less likely to occur if materials from several production days are incorporated into one lot for acceptance and payment determination. Combining material from more than one production day increases the chances that materials from more than one population will be combined into a joint population for the lot. This will tend to increase the variability associated with the combined population.

Typically, a lot is subdivided into equally sized sublots. This procedure promotes the use of stratified random sampling plans. Stratified sampling is used to ensure that the specimens for the sample are obtained from throughout the lot, and are not concentrated in one portion or section of the lot. Figure 24 illustrates the basic principle of stratified random sampling. The large rectangle represents a lot**,** perhaps one day's paving from which cores are to be obtained. Using a random selection process, it is possible (but not necessarily likely) that all of the cores could be selected within the first half of the lot. To avoid this possibility, the lot can be stratified into a number of sublots equal to the sample size to be selected from the lot. One core is then randomly selected from within each sublot. This ensures that each portion of the lot has the same chance of being selected while, at the same time, ensuring that the sample is spread out over the entire lot.

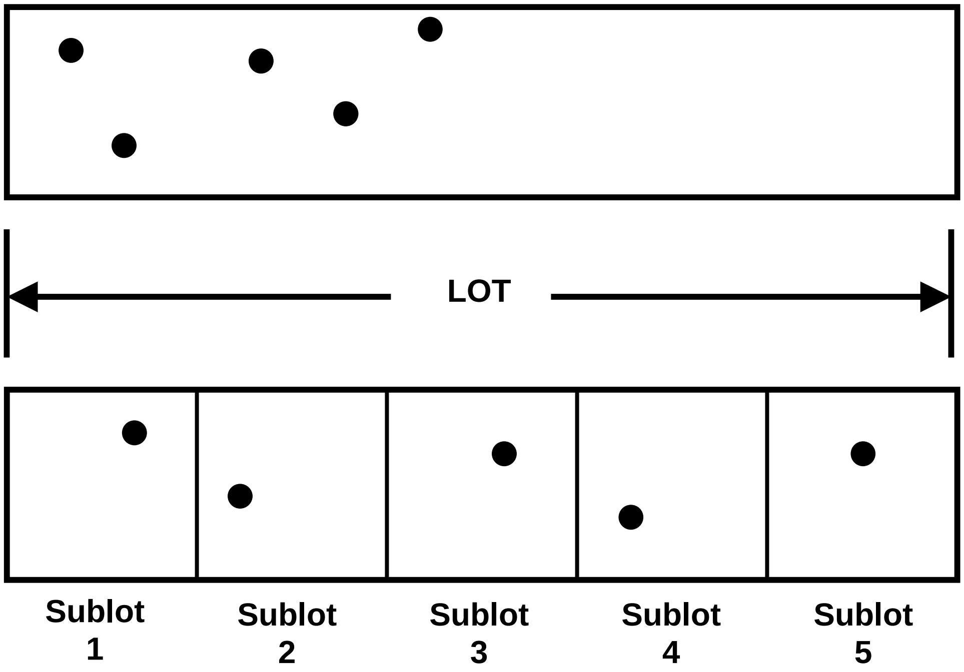


Figure 24. Examples of Random and Stratified Random Sampling

38.2. Selecting the Sample Size

The sample size is the number of test results used to judge the quality of a lot and, thus, is directly related to the lot size. One of the reasons to use larger lot sizes is the potential resultant increase in sample size. This tends to provide a lower level of risks, other considerations remaining equal. However, as noted above, a major assumption that is required is that all of the material and/or construction processes remain consistent throughout the total lot. This may be more difficult to achieve if the lot spans several production days, and obviously even more difficult to achieve if the entire project is used as the lot size.

From the above discussions it is obvious that there is a definite relationship between the lot size and sample size selections. Small lot sizes are not compatible with large sample sizes due to the large amount of testing that will be required. This can only work if the tests that will be used are nondestructive and can be completed quickly. This might be the case, for example, with the use of nuclear gauges for estimating density of HMAC pavements. Larger sample sizes can be used with large lot sizes to decrease risks of making incorrect decisions. However, the likelihood of combining materials from possibly different populations must be taken into consideration. For this to work, the variability data used to establish the AQL, RQL, and specification limits must also have been obtained from similar large lots.

In the prior discussions regarding establishing the appropriate process variability (see chapter 5), the importance was stressed of determining a process standard deviation that is consistent with the way in which a lot will be defined for acceptance. However, in practice the decision regarding lot size often cannot be determined with certainty early in the data collection process. It may be possible to determine whether or not the total project will be a lot, or whether acceptance will be on a lot–by–lot basis. In reality, the final decision regarding the sample size per lot cannot be made until an evaluation of the risks has been completed (see chapter 7).

If a major change in the definition of the lot, such as changing from lot–to–lot acceptance to project acceptance, is made after the typical process standard deviation has been calculated, then it will be necessary to re–evaluate the data to determine if a new typical process standard deviation should be used. If a revised standard deviation is selected, then it may be necessary to see if this will affect prior decisions regarding the AQL, RQL, and specification limits.

38.3. Sources of Information

The literature review suggested earlier in the manual should reveal the practices of other agencies. A 1998 TRB paper, entitled “Summary of Current QC/QA Practices for Hot-Mix Asphalt Concrete,” contains a survey of typical lot sizes, sample sizes, sample locations, etc., for HMAC. (21) Agencies can also be contacted directly to ascertain their experiences with the development of QA specifications.

38.4. Typical Lot Size, Sample Size, and Sampling Locations

Typical lot and sample sizes and sampling locations for payment determination are listed below for HMAC and PCC. The HMAC values are from an NCHRP project. (21) For screening tests, the tests are most often performed on a unit basis, similar to that used in attributes acceptance plans. The result is that acceptance is usually by the truckload or some other small discrete quantity with an accompanying small sample size.

For HMAC material properties: (21)

- The most often used lot sizes are 450 Mg to 4500 Mg, a day’s production, or a total item per mix design. Although some agencies may be using 450–Mg lots, more often the lot size will be about 1800 Mg. Some agencies reference the lot size to production rates. An example is the Virginia DOT, which uses an 1800–Mg lot unless normal daily production is in excess of 1800 Mg, at which time a 3600–Mg lot may be used.
- The most often used sublot sizes are 450 Mg to 900 Mg, 4 or 5 sublots/lot, and one test per 3 hours.
- The sampling location most often used is the truck, next is the mat. The choice of sample location is usually based on balancing the opportunity for obtaining samples quickly against sampling from a point at which the material must perform. If the contractor’s QC program is working properly, obtaining samples and quickly performing acceptance tests should not be as important an issue. However, in practice, most contractors want the

results of the acceptance tests in a timely manner to confirm their QC results. There may also be some concerns over the safety aspects of sampling from the mat.

For HMAC roadway compaction: (21)

- The most often used lot sizes are 360 Mg to 5400 Mg, a day's production, 300 meters to 1500 meters, or a total item per mix design. Again, 1814 Mg appears to be an often–used compromise.
- The most often used sublot sizes are 75 Mg to 1360 Mg, 1 to 5 tests per day, and 300 meters to 600 meters. The sublot size, and consequently the sample size, often depends on whether nuclear gauge readings or cores are used for acceptance. If nuclear gauges are used many more tests can be obtained in a reasonable amount of time than if cores are used. Also, the number of cores may be limited by the desire not to cut more samples from the pavement than absolutely necessary. Thus, the smaller sublot sizes reported are probably associated with the use of nuclear gauges and the larger sublot sizes are probably associated with the use of cores.
- The sampling location is the finished roadway using nuclear density gauges and/or cores.

For paving PCC:

- Often used lot sizes are a day's production, or 1.6 km.
- Often used sublot sizes are 4 hours, or 0.8 km.
- The sampling location may be the haul vehicle or the roadway.

For structural PCC:

- Often used lot sizes are a day's production or an item of construction.
- Often used sublot sizes are a stipulated quantity (number of trucks or cubic meters) or a stipulated time interval.

38.5. Closing: Lot Size and Sample Size

The discussion above is related to the initial establishment of lot sizes and sample sizes. These decisions are very likely to be economic ones, based on the resources available. These initial selections for lot size and sample sizes may need to be modified based on the analyses of risks that will be conducted next. The development of OC curves and their corresponding risk analyses are discussed in detail in the next chapter.

CHAPTER 7. EVALUATING RISKS

This chapter continues the discussion of Phase II of the specification development process. This chapter is intended to provide "how to use" best practices in evaluating the risks associated with the initial acceptance procedures that have been developed up to this point. The steps that are involved in this part of the process are identified in the flowchart in figure 25. The numbers in boxes before the titles of the following sections refer to the corresponding box in the flowchart.

39 Develop OC Curves and Evaluate Risks

Establishing the limits to be used for acceptance is an important step. Making the limits too restrictive deprives the contractor of a reasonable opportunity to meet the specification. Making them not sufficiently restrictive makes them ineffective in controlling quality. Selection of the limits relates to the determination of risks. The concept of risks for acceptance is similar to that discussed in chapter 5 for verification testing to evaluate whether test results came from the same population. The two types of risk discussed in chapter 5 are the seller's (or contractor's) risk, α, and the buyer's (or agency's) risk, β. The α risk is also called a Type I risk, and the β risk is also called a Type II risk. A well-written QA acceptance plan takes these risks into consideration in a manner that is fair to both the contractor and the agency. Too large a risk for either party undermines credibility.

39.1. Risks: Definitions and Concepts

39.1.1. Risks. Before proceeding further, some terms need to be formally defined. The TRB glossary [(2)] includes the following definitions:

- ***Seller's risk (α)****—also called* risk of a type I error. *The probability that an acceptance plan will erroneously reject acceptable quality level (AQL) material or construction with respect to a single acceptance quality characteristic. It is the risk the contractor or producer takes in having AQL material or construction rejected.*

- ***Buyer's risk (β)****—also called* risk of a type II error. *The probability that an acceptance plan will erroneously fully accept (100 percent or greater) rejectable quality level (RQL) material or construction with respect to a single acceptance quality characteristic. It is the risk the highway agency takes in having RQL material or construction fully accepted. [The probability of having RQL material or construction accepted (at any pay) may be considerably greater than the buyer's risk.]*

The α and β risk levels that might be appropriate vary depending upon the material or construction process that is involved. The appropriate risk level is a subjective decision that can vary from agency–to–agency. In reality, it is likely that few agencies have developed and evaluated the risk levels associated with their acceptance plans. While risk levels are an agency decision, AASHTO R–9, “Acceptance Sampling Plans for Highway Construction,” suggests the risk levels indicated in table 21.(22) It should be noted that large sample sizes, on the order of 10 to 20 or more, may be required to achieve some of the risk levels stipulated in this table.

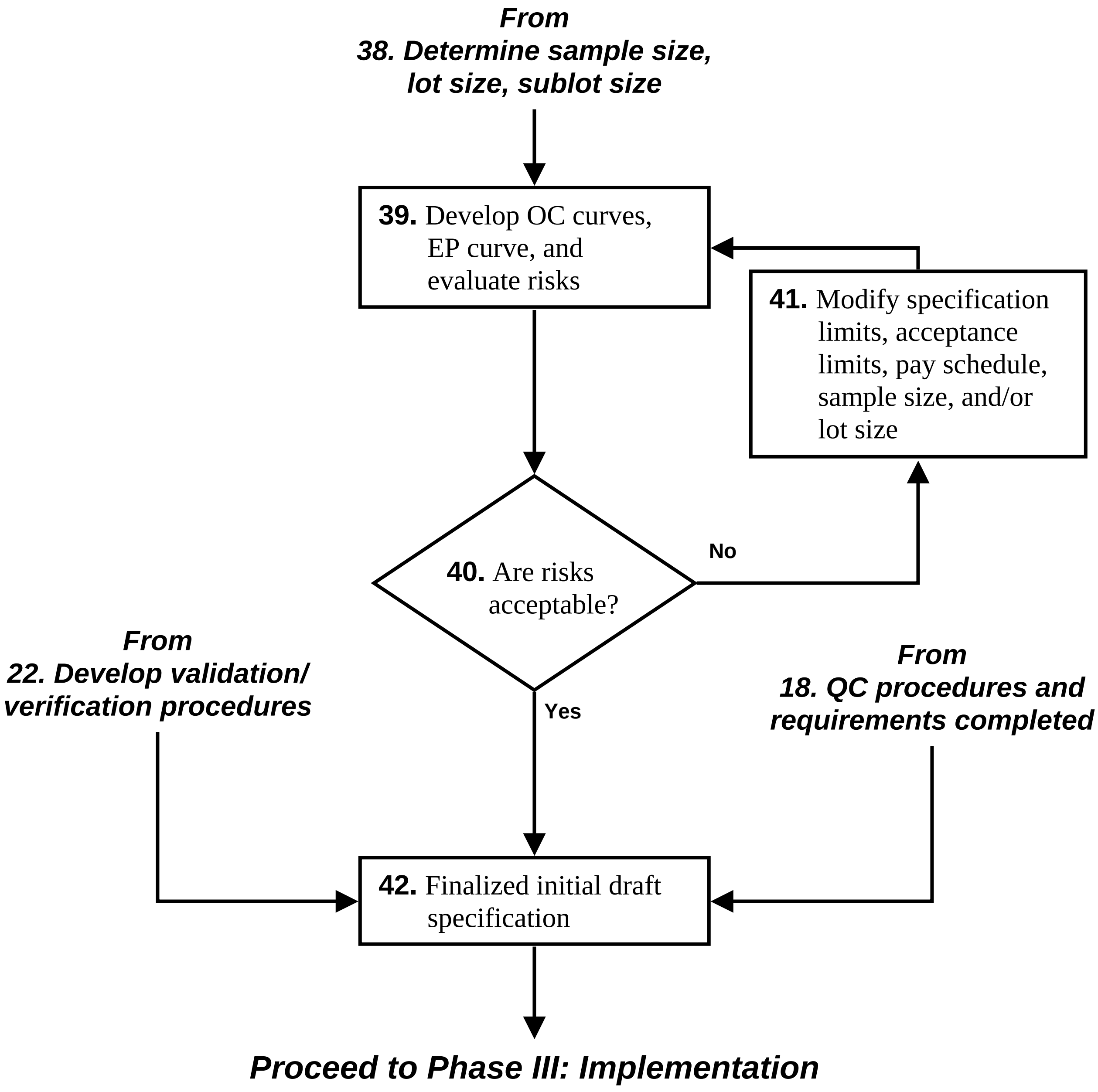


Figure 25. Flowchart for Risk Analysis Portion of Phase II

Table 21. Risk Levels Suggested in AASHTO R–9 (22)

Criticality [1]	Recommended α	Recommended β
Critical	0.050	0.005
Major	0.010	0.050
Minor	0.005	0.100
Contractual	0.001	0.200

[1] Critical: when the requirement is essential to preservation of life.
Major: when the requirement is necessary for the prevention of substantial financial loss.
Minor: when the requirement does not materially affect performance.
Contractual: when the requirement is established only to provide uniform standards for bidding.

As noted in the section on verification testing in chapter 5, the concept of α and β risks derives from statistical hypothesis testing where there is either a right or wrong decision. As such, when α and β risks are applied to materials or construction they are only truly appropriate for the case of a pass/fail or accept/reject decision and, in fact, may lead to considerable confusion if an attempt is made to apply them to the payment adjustment case. When materials not only can be accepted or rejected, but can also be accepted at an adjusted payment, then additional interpretations or clarifications must be applied to the definitions for these risks in an effort to manipulate them to apply to the payment adjustment situation.

For example, in the definition for buyer's risk above, it states that β is the probability that RQL material will be accepted at 100 percent payment or greater. The definition must then go on to point out that there is a much greater probability that the RQL material will receive some reduced payment. While it is not stated as directly, the same reasoning is true for the seller's risk. The definition indicates that α is the probability that AQL material will be rejected. Although not stated in the definition, it is also true that there is a much greater probability that the AQL material will be accepted at a reduced payment.

39.1.2. OC Curves. The buyer's and seller's risks are very narrowly defined to occur at only two specific quality levels. The buyer's risk is the probability of accepting material that is exactly at the RQL level of quality, while the seller's risk is the probability of rejecting material that is exactly at the AQL level of quality. These definitions do not therefore provide a very good indication of the risks over a wide range of possible quality levels. To evaluate how the acceptance plan will actually perform in practice, it is necessary to construct an OC curve. The TRB glossary (2) includes the following definition:

- ***OC curve***—*A graphic representation of an acceptance plan that shows the relationship between the actual quality of a lot and either (1) the probability of its acceptance (for* *accept/reject acceptance plans) or (2) the probability of its acceptance at various payment levels (for acceptance plans that include pay adjustment provisions).*

An example of an OC curve for a pass/fail or accept/reject acceptance plan, case (a) in the above definition, is shown in figure 26. Probability of acceptance is shown on the vertical axis for the range of quality levels indicated on the horizontal axis. An example of an OC curve for an acceptance plan with payment adjustment provisions, case (b) in the above definition, is shown in figure 27. The axes are the same as for figure 26, but there are multiple curves, one for each of several selected payment levels, plotted.

Each curve plotted in figure 27 represents the probability of receiving a payment factor equal to or greater than the one indicated for the line. For example, for the OC curves in figure 27, material that is of exactly AQL quality has approximately a 45 percent chance of receiving a payment factor of 1.04 (104 percent) or greater. This same AQL material has approximately a 55 percent chance of receiving full payment (100 percent) or greater, which also means that it has approximately a 45 percent chance of receiving less than 100 percent payment. This AQL material has essentially a 100 percent chance of receiving a payment factor of 0.80 (80 percent) or greater.

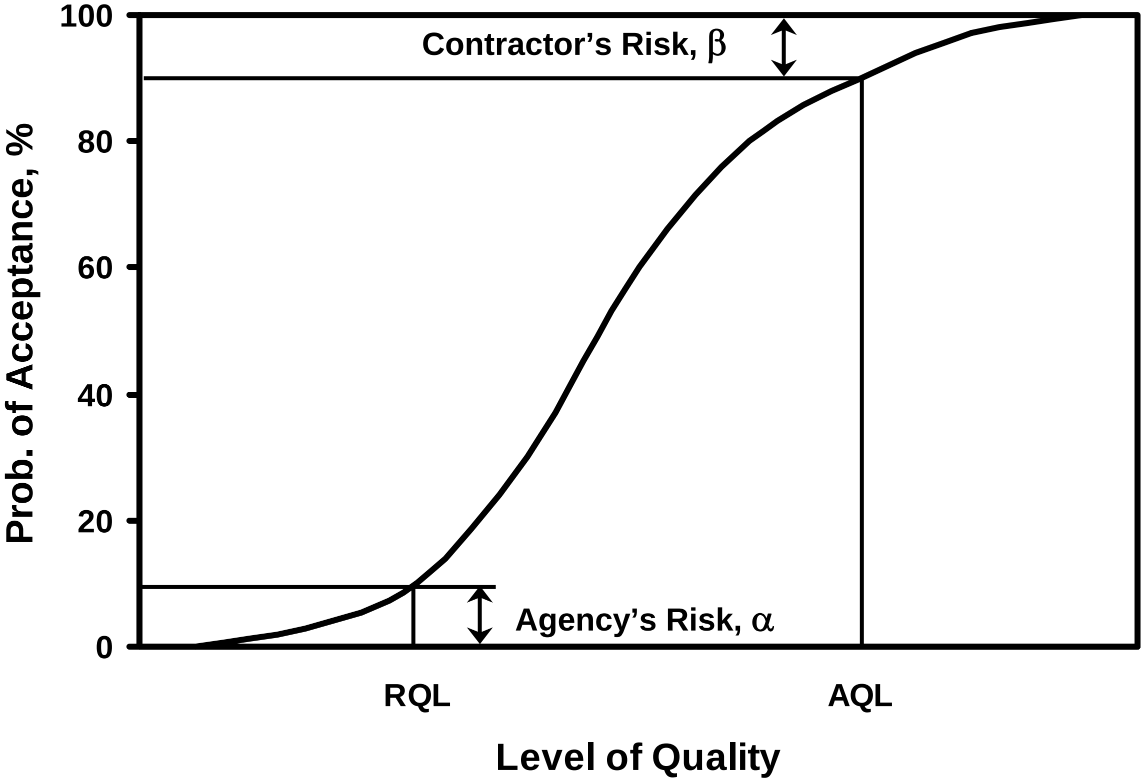


Figure 26. Typical OC Curve for an Accept/Reject Acceptance Plan

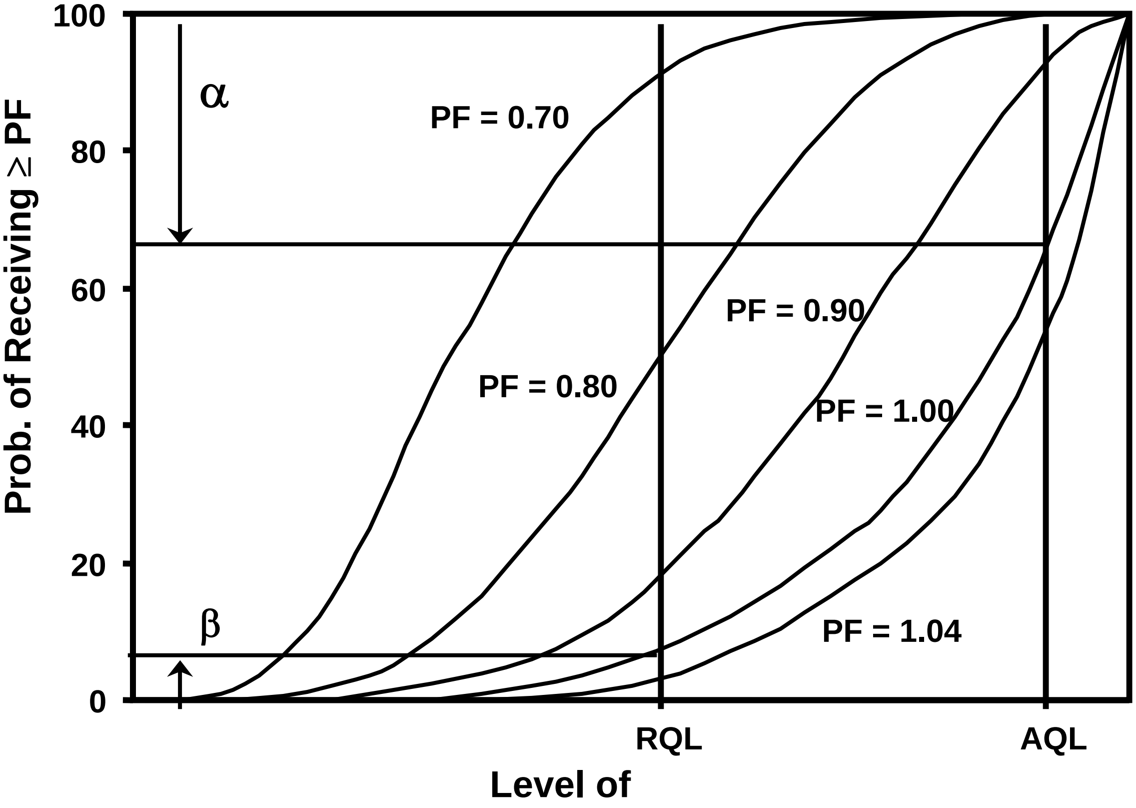


Figure 27. Typical OC Curves for an Acceptance Plan with Payment Adjustments

On the other hand, for the OC curves in figure 27, material that is of exactly RQL quality has approximately a 30 percent chance of receiving a payment factor of 0.80 (80 percent) or greater, and nearly an 80 percent chance of receiving a payment factor of 0.70 (70 percent) or greater. Similar payment probabilities can be determined for any level of actual quality, and additional curves could be developed for any specific value of payment factor.

39.1.3. Expected Payment Curves. Figure 27 clearly shows that consideration of only α and β risks is clearly not sufficient when payment adjustments are used. From figure 27 it can also be seen that using multiple OC curves is not an easy way to evaluate an acceptance plan. It would be convenient to have a single curve that can represent the operation of the plan as opposed to many different curves for each plan. Another way to present the payment performance for an acceptance plan is with what is call an expected payment (EP) curve. The TRB glossary (2) includes the following definition:

- ***EP curve****—A graphic representation of an acceptance plan that shows the relation between the actual quality of a lot and its EP (i.e., mathematical pay expectation, or the average pay the contractor can expect to receive over the long run for submitted lots of a* *given quality). [Both OC and EP curves should be used to evaluate how well an acceptance plan is theoretically expected to work.]*

An example of an EP curve is shown in figure 28. Quality levels are indicated on the horizontal axis in the usual manner, but instead of probability of acceptance, the vertical axis gives the expected (long–term average) payment factor as a percent of the contract price.

Although the risks have a different interpretation when associated with EP curves than with OC curves, the same type of information is provided. For the example in figure 28, AQL work receives an expected payment of 100 percent, as desired, while truly superior work that is better than the AQL receives an expected payment of 102 percent. At the other extreme, RQL work corresponds to an expected payment of 70 percent. For still lower levels of quality, the curve levels off at a minimum expected payment of 50 percent.

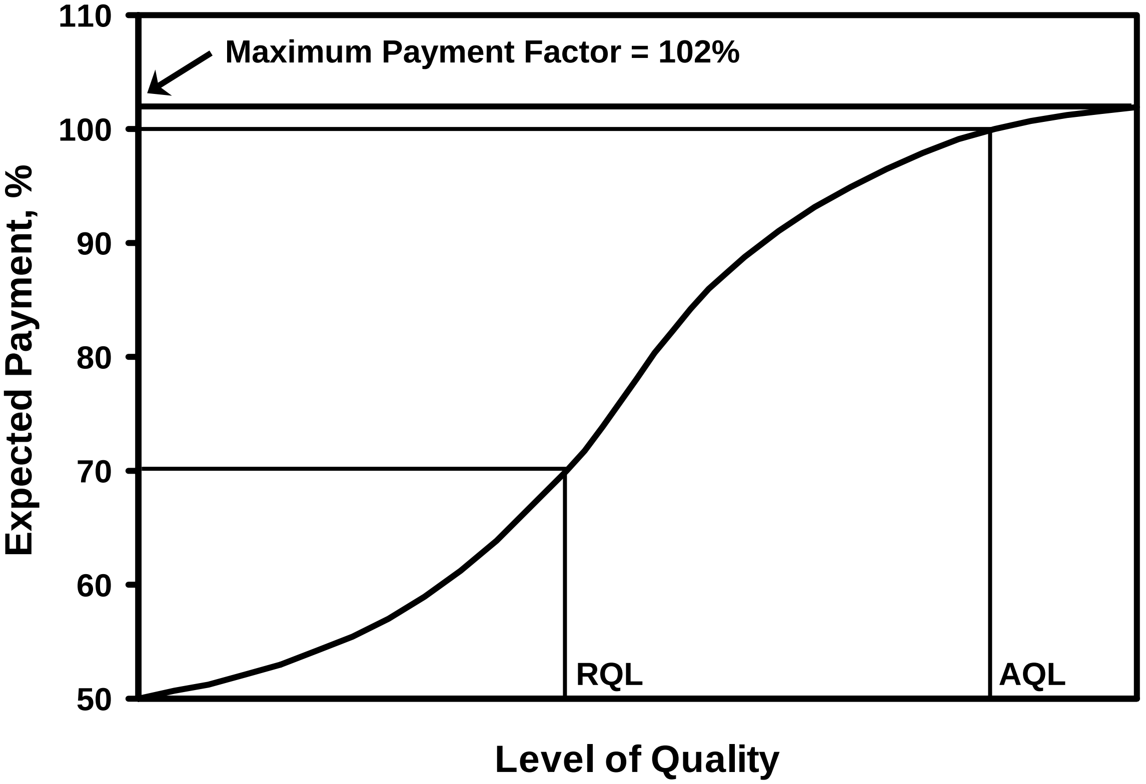


Figure 28. Typical EP Curve

Simplified Example: α and β Risks and an OC Curve

A simplified example of how risks are related to specification limits can be illustrated by considering primitive acceptance plans that were based on measuring and accepting a property based on only one test. Suppose that an accept/reject acceptance plan for asphalt content has been developed based on the definitions for AQL material and RQL material that follow.

Define AQL Material. It is assumed that asphalt content follows a normal distribution. It has been determined that for asphalt content, acceptable material has a standard deviation of about 0.20 percent when the mean is close to the target JMF value. If the JMF has established the target as 6.0 percent asphalt content, the AQL is therefore a lot (population) with a mean of 6.0 percent and a standard deviation of 0.20 percent. Figure 29 shows an AQL population.

Define RQL Material. Additionally, unacceptable material might be defined as that for which the mean differs from the target value by 0.4 percent or more, as long as the standard deviation does not exceed 0.20 percent. (Other definitions would be equally valid.) The RQL is therefore a lot (population) with a standard deviation of 0.20 percent and a mean of 5.6 percent or lower, or 6.4 percent or higher. Examples of RQL populations are shown in figure 30.

Determine α Risk. Suppose the agency has established the specification limits, i.e., the limits within which individual asphalt content results must fall, as the JMF ± 0.40. For a JMF target value of 6.0 percent, this establishes the specification limits as 5.60 percent and 6.40 percent. An AQL population is shown along with the specification limits in figure 31. The α risk is the probability that a single test result from this AQL lot would be outside of the allowable specification range of 5.60 percent to 6.40 percent. This is the α risk to the contractor because if a test falls outside these limits the agency will erroneously reject the material. The *Z*–statistics can be calculated and used in conjunction with the standard normal distribution table (table 7) to determine this probability to be 0.0456 or 4.56 percent.

Determine β Risk. Figure 32 shows an RQL population with its mean at 5.60 percent and standard deviation of 0.20 percent. The RQL population could also have its mean at 6.40 percent. The RQL population can either be too low or too high, but not both at the same time. The β risk is the probability that a single test result from this RQL lot would be within of the allowable specification range of 5.60 percent to 6.40 percent. This is the risk to the agency because if a test result falls in this range, the agency will erroneously accept the material. From figure 32 this probability can be seen to be 0.50 or 50 percent.

Develop the OC Curve. Similarly, the probabilities of acceptance for lots with means of any value, e.g., 5.20 percent, 5.40 percent, 5.60 percent, etc., can be calculated and plotted to form the OC curve shown in figure 33. The AQL and RQL are also noted on the figure. It should be noted that it is purely coincidental that the OC curve in figure 33 has the appearance of a normal curve.

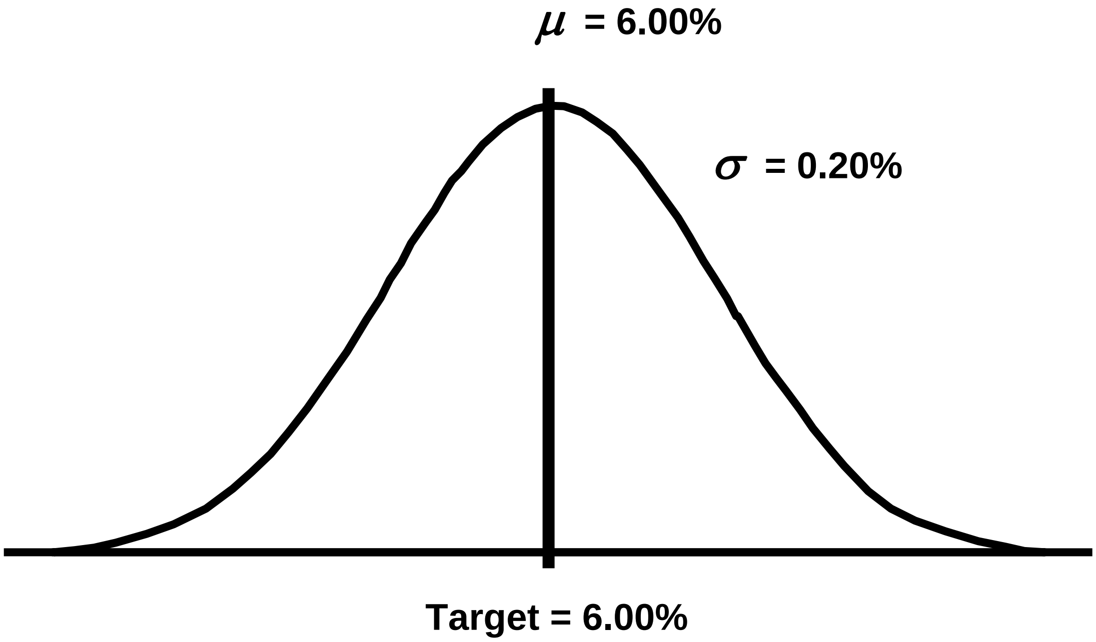


Figure 29. AQL Population for Simplified α and β Risks Example

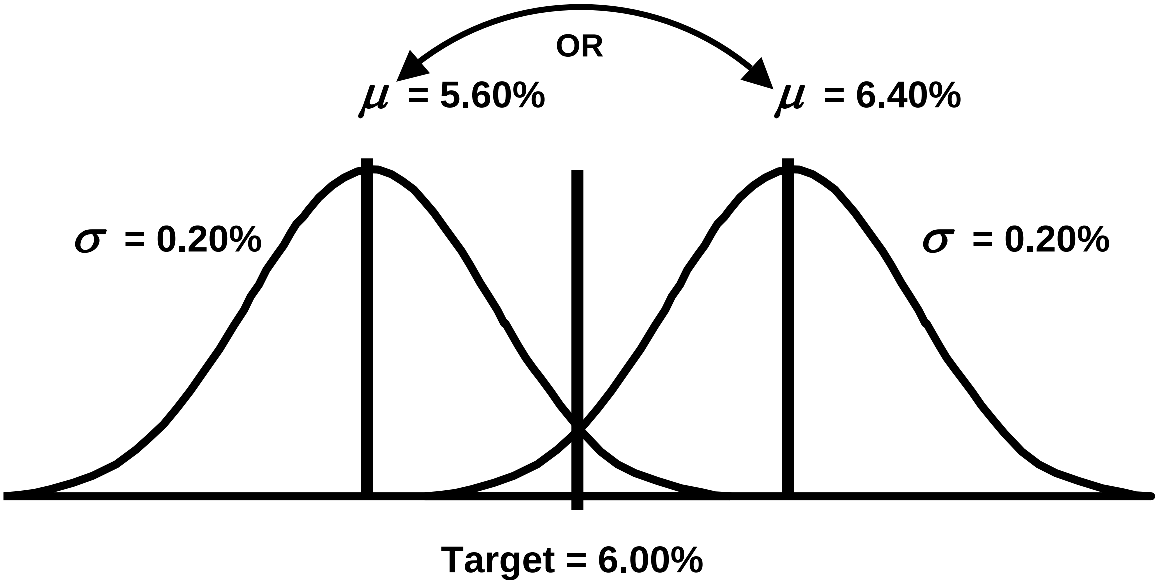


Figure 30. RQL Populations for Simplified α and β Risks Example

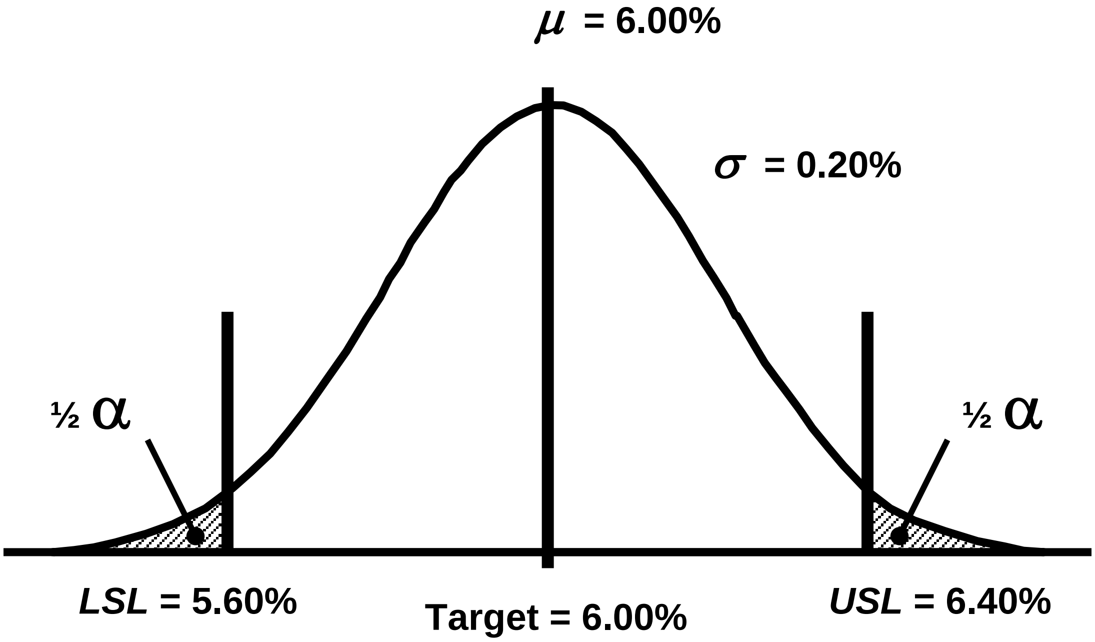


Figure 31. Illustration of the α Risk for Simplified Example

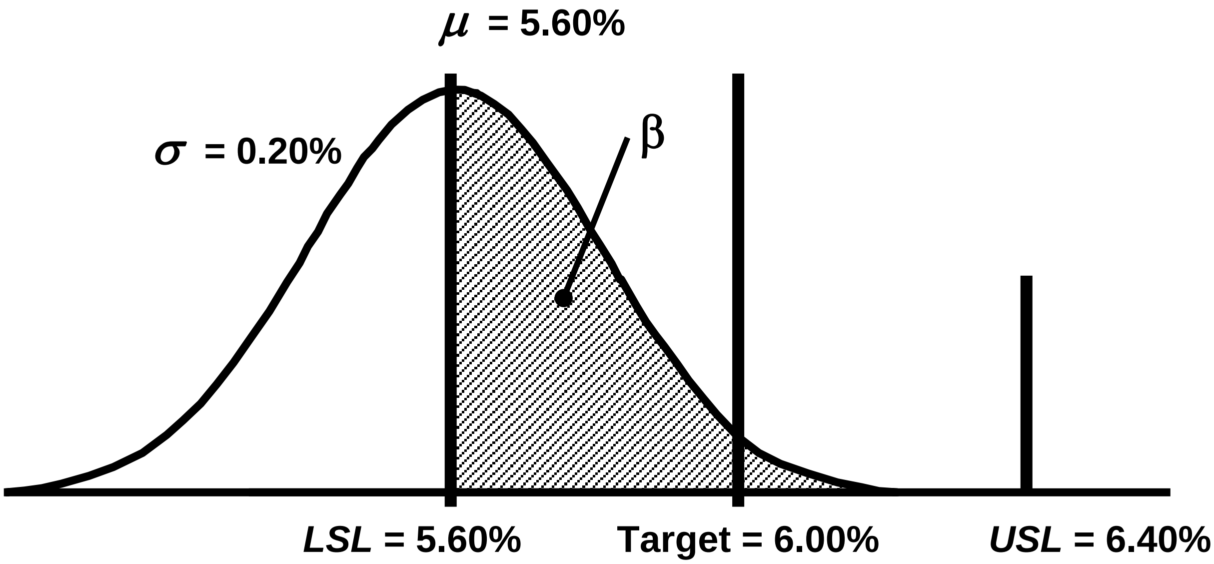


Figure 32. Illustration of the β Risk for Simplified Example

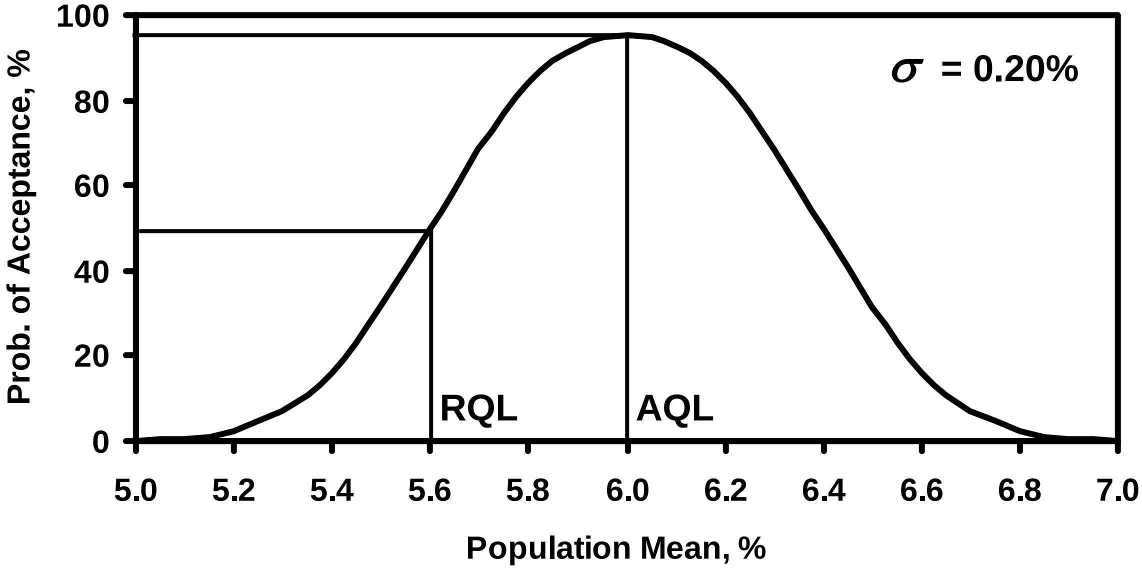


Figure 33. OC Curve for Simplified α and β Risks Example

39.2. OC Curves for PWL or PD Acceptance Plans

As with any acceptance plan that bases the acceptance decision on a sample, there are risks associated with PWL or PD acceptance plans. The above example demonstrated the calculation of risks for a simple acceptance plan based on an assumed known standard deviation, and with the acceptance decision based on only a single test result. The risks associated with PWL or PD acceptance plans cannot be calculated so easily.

For PWL or PD acceptance plans the risks are almost always determined by means of computer simulation. It is, however, possible to illustrate the risks associated with using a sample to estimate PWL by means of a simplified attributes example.

Simplified Example

Assume that we have a bag that has 100 marbles. Further assume that the bag has 70 white marbles and 30 blue marbles. Also assume that we wish to take a sample of 10 marbles to estimate the percentage of the marbles in the bag that are blue.

It is easy to estimate the percentage of blue marbles from a sample of 10 marbles. However, each sample of 10 marbles will not yield the same percentage of blue marbles. The first sample of 10 marbles might contain 3 blue marbles, thereby yielding an estimate of 30 percent blue marbles. However, it could also have only one blue marble, or five blue marbles. In each of these cases the estimate from the sample will be fairly far from the true value of 30 percent.

The histogram in figure 34 shows the results of 100 samples, each with 10 marbles. While the individual sample results could be quite far from the actual percentage in the population, the average of the 100 samples is quite close to the true population value. Also, most of the sample values are close to the actual population percentage, with fewer values as the estimate becomes farther from the actual population percentage. Although simplified, this example clearly shows how the PWL values estimated from samples can vary. The long–run average of the sample averages will tend to equal the true population PWL value, but there is a risk that any individual estimate may either over–estimate or underestimate the true population PWL value.

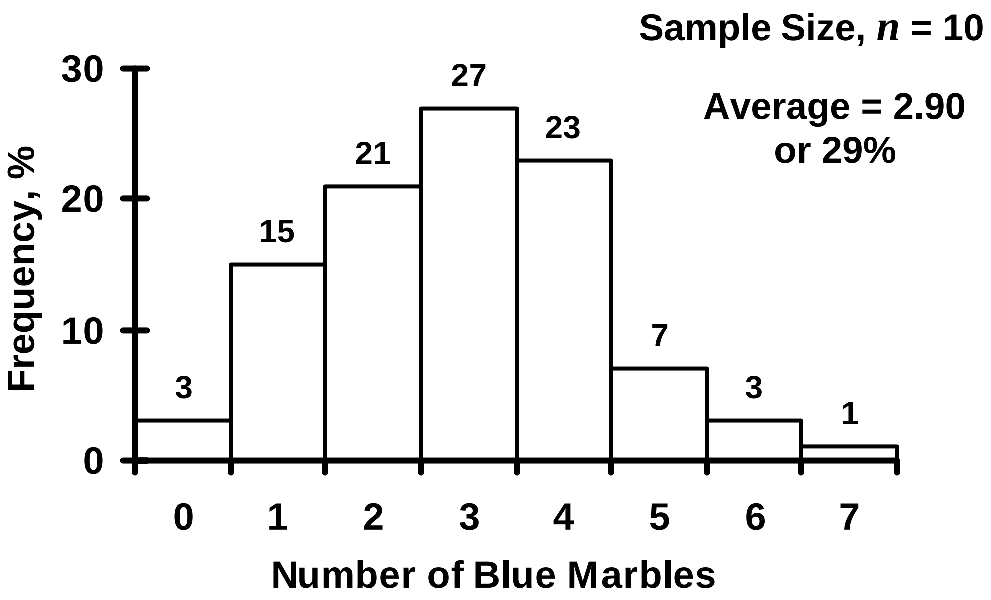


Figure 34. Histogram of PWL Estimates for Simplified PWL Example

39.2.1. Computer Simulation. As noted above, calculating the risks for actual PWL acceptance plans is much more involved than the simplified example from figure 34. In fact, computer simulation is almost always used to develop α and β risks, as well as OC and EP curves. OCPLOT, a user–friendly program that develops OC and EP curves by computer simulation, was developed as part of FHWA Demonstration Project No. 89. This program is explained in detail in the report for that project,(18) and is also presented in appendix M along with some examples.

OCPLOT can be used to develop OC curves for accept/reject acceptance plans. It can also be used with a stipulated payment equation to determine the probability of receiving a lot payment factor greater than or equal to any specific value. In this way, it can be used to plot multiple OC curves similar to those in figure 27. The program can also develop EP curves for a given payment equation. The program is also capable of simulating acceptance plans containing retest provisions.

39.3. Evaluating the Risks

39.3.1. Accept/Reject Acceptance Plans. How potential risks are evaluated depends upon the type of acceptance plan that is used. The evaluation of risks is rather straightforward for accept/reject (pass/fail) acceptance plans. As noted above, α and β risks and OC curves were developed specifically for this type of situation. Therefore, they can be directly used to assess the risks to both parties.

To reiterate, the α risk is the probability that AQL material will be rejected; while the β risk is the probability that RQL material will be accepted. However, since contractors will not operate at only these two quality levels, to fully consider risks the OC curve, which illustrates probability of acceptance for any quality level, must be developed for the acceptance plan under consideration. An example will help to illustrate how this can be done.

Example: Accept/Reject Acceptance Plans–OC Curves

The previously discussed OCPLOT program can be used to determine the α and β risks and to plot the OC curve for a sample acceptance plan. Suppose that an agency decides to use asphalt content as an accept/reject property for an HMAC pavement (note, this is not recommended, but is used here solely for the purpose of illustrating the use of an OC curve for an accept/reject situation). Further suppose that the agency has established for asphalt content a lower specification limit of 5.60 percent and an upper specification limit of 6.40 percent. The agency has decided to use the PWL, based on a sample of size 4, as the quality measure. The agency has selected 90 PWL for the AQL and 50 PWL for the RQL. The lot will be accepted if the estimated PWL is greater than or equal to 70. Table 22 and figure 35 show the results of the OCPLOT analysis of this proposed acceptance plan.

From table 22 it can be seen that the seller's risk is α = 1.000 – 0.905 = 0.095 (or 9.5 percent) and the buyer's risk is β = 0.144 (or 14.4 percent). Further, both table 22 and figure 35 show the probability of acceptance over the total range of possible lot quality levels, as defined by the actual PWL for the lot. The agency would need to decide whether or not it considers these levels of risk to be appropriate.

39.3.2. Payment Adjustment Acceptance Plans. The evaluation of risks becomes much more complicated when the acceptance plan includes payment adjustment provisions. The concepts of α and β risks, which were developed from hypothesis testing where there is a yes or no decision, i.e., reject or fail to reject the null hypothesis, are not sufficient when the decision involves not only accept or reject, but also accept at an adjusted payment level.

Table 22. OC Table from OCPLOT for the Example Problem

Population PWL	Probability of Acceptance
100	1.000
95	0.976
90 (AQL)	0.905 (α = 0.095)
85	0.810
80	0.696
75	0.579
70	0.466
65	0.363
60	0.288
55	0.200
50 (RQL)	0.144 (β = 0.144)
45	0.093
40	0.066
35	0.038
30	0.021
25	0.013
20	0.000

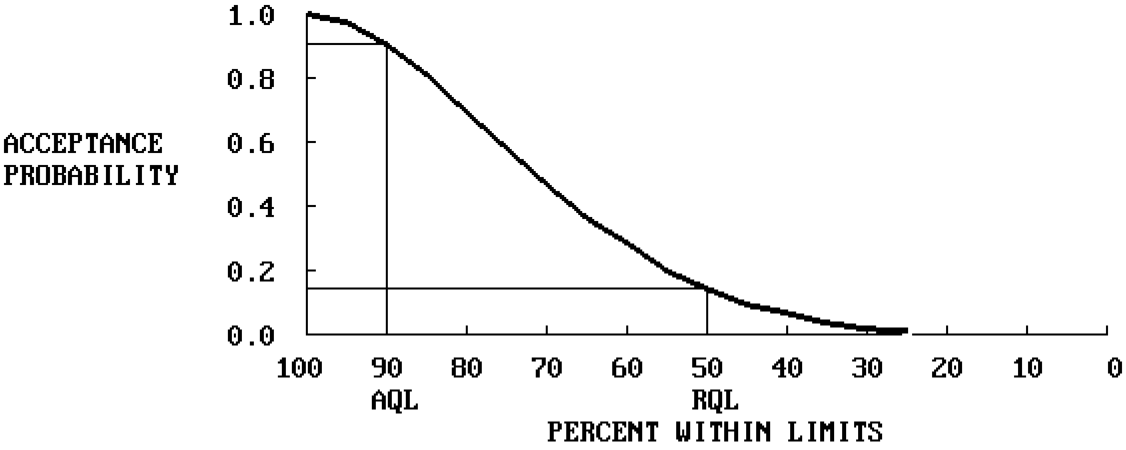


Figure 35. OC Curve from OCPLOT for the Accept/Reject Acceptance Plan Example

The TRB glossary (2) definitions for seller's and buyer's risks that are presented above do not attempt to incorporate the concept of payment adjustments. The seller's risk is defined as the *probability that an acceptance plan will erroneously reject AQL material or construction.* This disregards the fact that the material or construction can be accepted at full payment, increased payment, or decreased payment. In other words, whether or not a lot received 105 percent, 100 percent, 75 percent, or 50 percent payment would have no impact with regard to the seller's risk based on this definition. Obviously, however, these different payment levels would have quite an impact on how the contractor perceived its risks.

Similarly, the buyer's risk is defined as the *probability that an acceptance plan will erroneously fully accept (100 percent or greater) RQL material or construction.* Once again, this definition disregards the impact of partial payments when determining the buyer's risk. However, when considering its risks the agency will certainly be interested in the probability of accepting RQL material at reduced payment levels as well as at 100 percent payment or greater.

The use of α and β risks to evaluate payment adjustment acceptance plans is simply not sufficient. Some additional method or methods is/are necessary to properly evaluate the risks when payment adjustments are added to the acceptance decision options. The expected payment, or EP, (see figure 28) is another method for considering the payment adjustment aspects of the acceptance plan. However, EP alone is also not sufficient to fully evaluate the risks that are involved. Multiple OC curves for various payment levels (see figure 27) should also be developed when evaluating acceptance plans with payment adjustment provisions. An example will help to illustrate the evaluation of risks for payment adjustment acceptance plans.

Example: Payment Adjustment Acceptance Plans—EP Curves

Consider the previous asphalt content example where the sample size was 4, the allowable specification range was 5.60 percent to 6.40 percent, and the AQL and RQL were defined as 90 PWL and 50 PWL, respectively. However, instead of a simple accept/reject acceptance plan, the agency chooses to use equation 28 to establish the payment factor for a lot:

$$PF = 55 + (0.50 \times \mathrm{PWL}) \tag{28}$$

where: PF = payment factor for the lot, as a percent of contract price.
PWL = estimated PWL value for the lot.

From the above equation, it can be seen that the maximum payment factor is 105 percent at 100 PWL, while the payment factor at the AQL will be 100 percent and the payment factor at the RQL will be 80 percent. It is generally accepted that the average payment for AQL material should be 100 percent. In this example, the payment factor at the AQL is 100 percent, exactly as intended. However, if the payment equation is not developed properly, the average payment factor may turn out to be above or below 100 percent at the AQL. If this is the case, the agency should determine if an expected payment other than 100 percent for AQL material is acceptable.

With the above information, the OCPLOT program can be used to develop the EP curve shown in figure 36. It can be seen in this figure that, as desired, the EP for AQL material is 100 percent. This means that a contractor that consistently produces material that just meets the minimum requirements, i.e., AQL material, will receive an average payment factor of 100 percent in the long–run. Similarly, the EP for RQL material is 80 percent as desired from the payment equation.

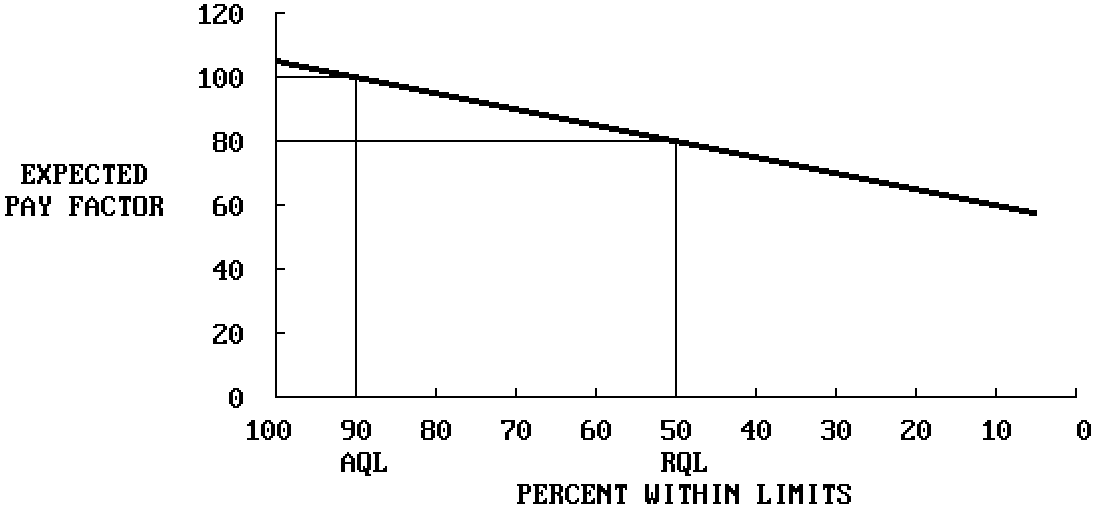


Figure 36. EP Curve from OCPLOT for the Payment Adjustment Acceptance Plan Example

The EP curve has the advantage of combining all of the possible payment levels into a single expected, or long–term average, payment for each given level of quality. While it is a major improvement over only considering α and β risks, the use of the EP alone still has some serious deficiencies. The primary deficiency in the use of EP alone is that, while it considers the average long–term payment factor, it fails to consider for a given quality level the variability of the individual lot payment factors that comprise this long–term average. This variability is directly related to the sample size. That is, the variability about the average payment factor decreases as the size of the individual samples increases. To fully evaluate the risks it is necessary to also consider this variability about the expected payment values.

The OCPLOT program output can be used to demonstrate this variability of the individual lot payment factors. Figure 37 shows for an AQL population a histogram that displays the individual lot PWL estimates along with their corresponding payment values for 1,000 simulated lots using a sample of size of 4 for each individual lot. Figure 38 shows similar information for an RQL population. The high degree of variability of the individual lot payment factors is obvious from this histogram. However, over a large number of lots, the high and low estimates for lot PWL will tend to balance out to give the correct average payment factor.

If, however, there are only a small number of lots on a project, then it will be possible that a significantly low estimated PWL value could negatively impact the payment that the contractor should have received. Similarly, larger PWL estimates could be obtained that would provide a larger payment than is deserved. A contractor would be wise to target a quality level above the AQL, particularly on smaller projects, to ensure that this variability of individual lot PWL estimates does not create a problem. However, as discussed elsewhere in this manual, it is often the practice of contractors to bid projects with the anticipation of receiving the maximum incentive payments. If this is the case, it is unlikely that contractors will target their processes at the AQL. It is more likely that they will target their processes for greater than AQL quality to try to maximize their incentive payment. In this event, the variability of the individual lot payment factors will not likely pose a serious problem to either the contractor or the agency.

The variability associated with the estimate of the lot PWL can be reduced by increasing the size of the sample obtained from each lot. Figure 39 shows a histogram that displays for an AQL population the individual lot PWL estimates along with their corresponding payment values for 1,000 simulated lots using a sample of size of 20 for each individual lot. Figure 40 shows similar information for an RQL population. When these figures are compared with figures 37 and 38, for samples of size 4, the smaller spread of the individual PWL and payment factor estimates is apparent.

Even if it reduces the variability of the PWL estimates, and hence the risk to both the contractor and the agency, it may not be practical or economical to use large sample sizes unless correspondingly large lot sizes are also used. The use of a very large lot, possibly even the total project, will allow larger sample sizes, but also introduces problems of its own. As noted in chapter 6, a major assumption that is required is that all of the material and/or construction processes remain consistent throughout the total lot.

Over the course of a long project, changes in weather, materials, rolling patterns, mix designs, etc., are likely to lead to variations throughout the project. Combining all of these together may result in a normal distribution, albeit one with a larger variability than the individual production lots, but this may not be the best method to evaluate a project. If there are “bad” segments on a project, it might be better to see them penalized on a lot–by–lot basis than to have them lumped together with the “good” material from all of the other lots.

While figures 37 through 40 clearly illustrate the relative variabilities of the individual PWL and payment factor estimates associated with different sample sizes, they do not provide any quantitative measure for the variabilities. One way to quantify these variabilities would be to calculate the standard deviation of the individual PWL or payment estimates. This is not discussed in this manual, but is presented and discussed in the technical report for this project.(17)

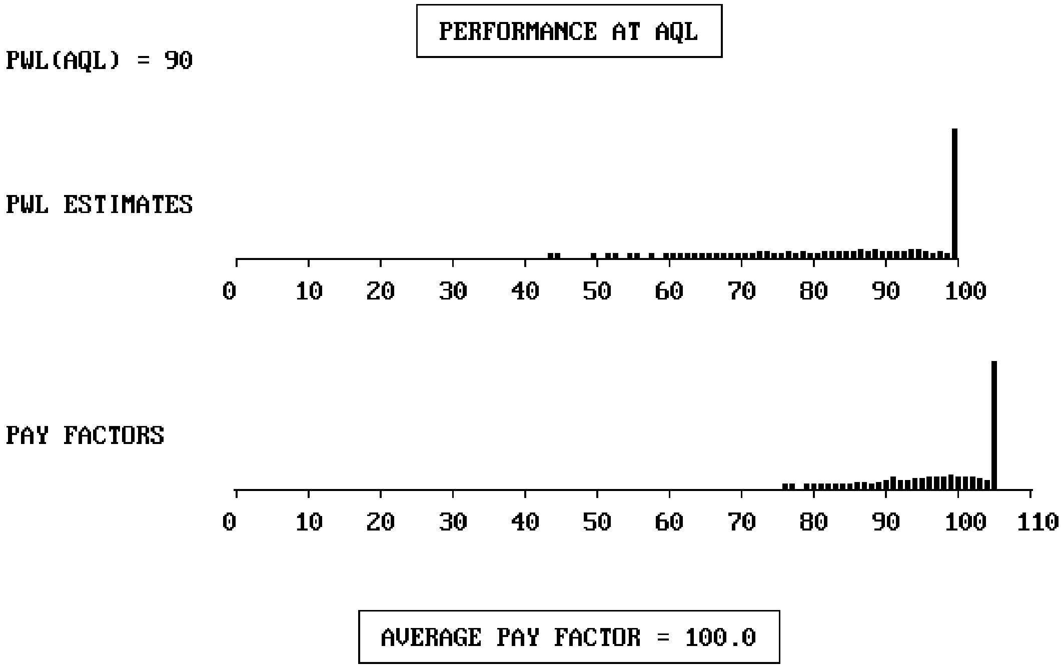


Figure 37. Histogram for an AQL Population Showing Variability of Individual PWL and Payment Factor Estimates for a Sample Size of 4

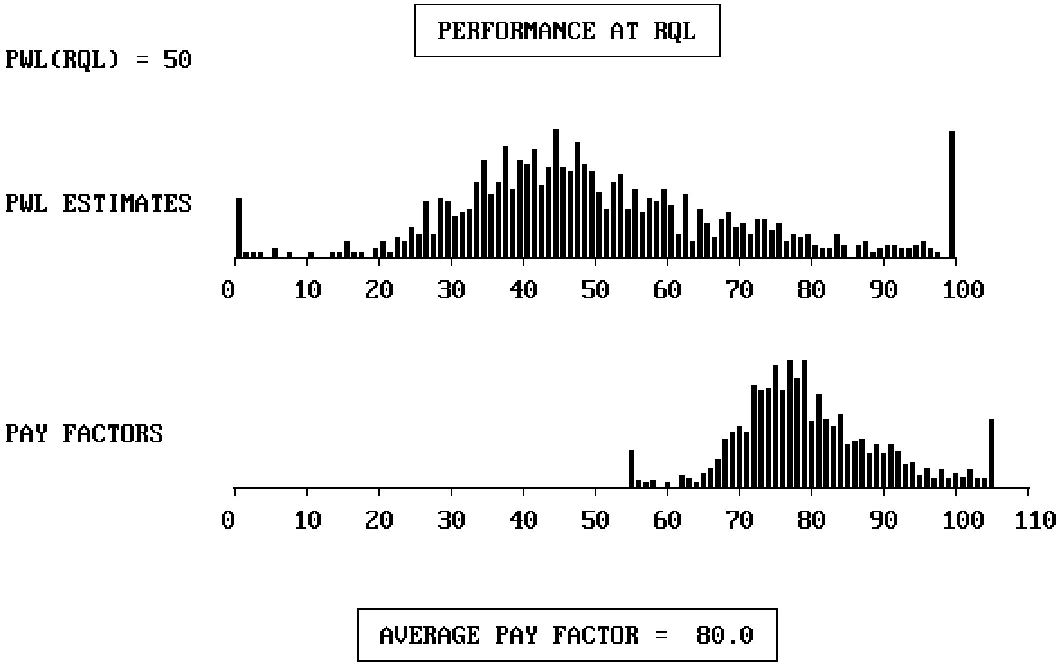


Figure 38. Histogram for an RQL Population Showing Variability of Individual PWL and Payment Factor Estimates for a Sample Size of 4

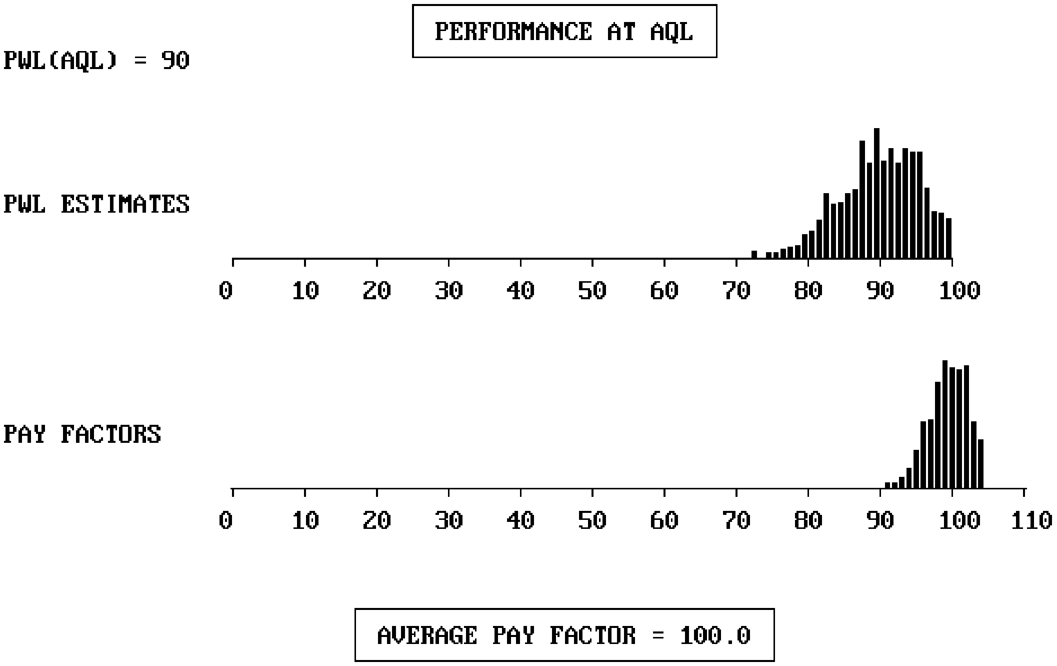


Figure 39. Histogram for an AQL Population Showing Variability of Individual PWL and Payment Factor Estimates for a Sample Size of 20

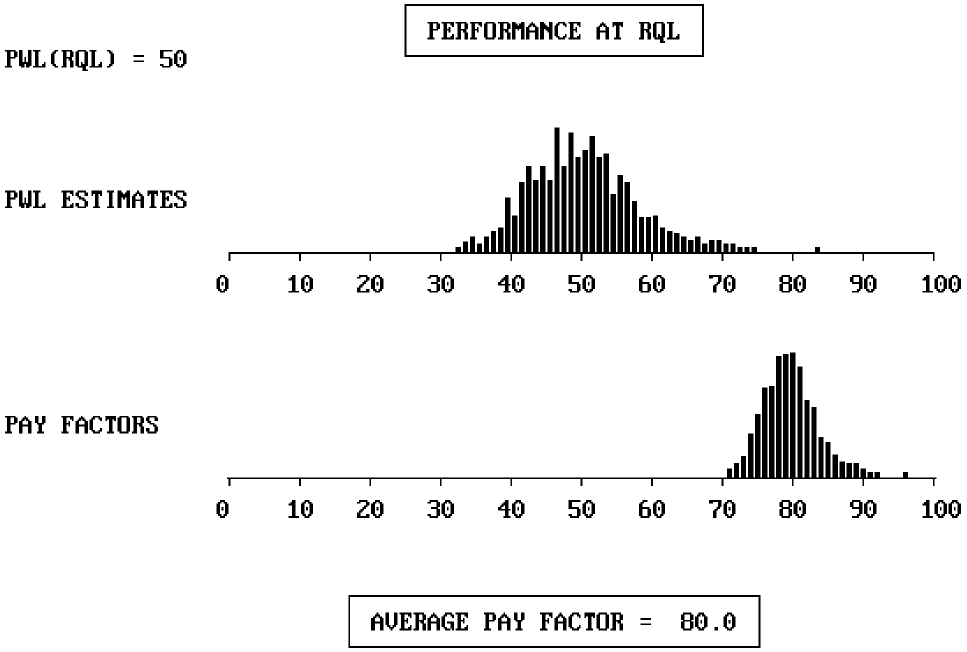


Figure 40. Histogram for an RQL Population Showing Variability of Individual PWL and Payment Factor Estimates for a Sample Size of 20

Example: Payment Adjustment Acceptance Plans—Multiple OC Curves

Another step that is necessary to evaluate fully the risks for a payment adjustment acceptance plan is to plot OC curves, such as those shown in figure 27, associated with receiving various payment factors. As shown in appendix M, the OCPLOT program can be used to develop these curves, although each curve must be developed individually and then manually combined onto a single set of axes.

Suppose that the OCPLOT program is used to develop multiple OC curves for the asphalt content acceptance plan from the previous example. Figure 41 shows OC curves for the probability of receiving greater than or equal to various levels of payment factor for a sample of size of 4 using the payment relationship shown in equation 28. These OC curves would be considered along with the EP curve from the previous example to evaluate the risks associated with the acceptance plan.

While the EP curve in figure 36 shows that the average long–term payment is 100 percent for AQL material, the OC curves in figure 41 show that the probability is less than 60 percent that any individual lot of AQL material will receive 100 percent payment or greater. This means that there is nearly a 40 percent chance that a contractor would receive less than full payment for a lot that was of AQL quality. This risk, which would be considered to be α (if α is defined as the probability that AQL material will receive less than full payment), seems high. However, it is somewhat offset by the fact that the OC curves also indicate that there is over a 40 percent chance of receiving a payment of 104 percent or greater.

40

Determine Whether or Not the Risks Are Acceptable

The OC curves and EP curves describe the operation of the acceptance plan such that the risks can be evaluated throughout the entire quality regime. If the risks are considered acceptable, no modifications to the initial acceptance plan are necessary. However, if the risks are considered unacceptable in terms of being too high for both or either party, a reassessment of the acceptance plan is necessary.

As shown in the previous section, there is no easy answer to the question “Are the risks acceptable?” since this is to a great extent a subject of opinion, and opinions may vary from agency–to–agency. Table 21 can provide some guidance regarding α and β risk levels, but these risks are not very useful when price adjustment acceptance plans are used. Even in accept/reject acceptance plans, the α and β risks apply at only two specific levels of quality. An OC curve is still necessary to evaluate the risks over the full range of possible quality levels.

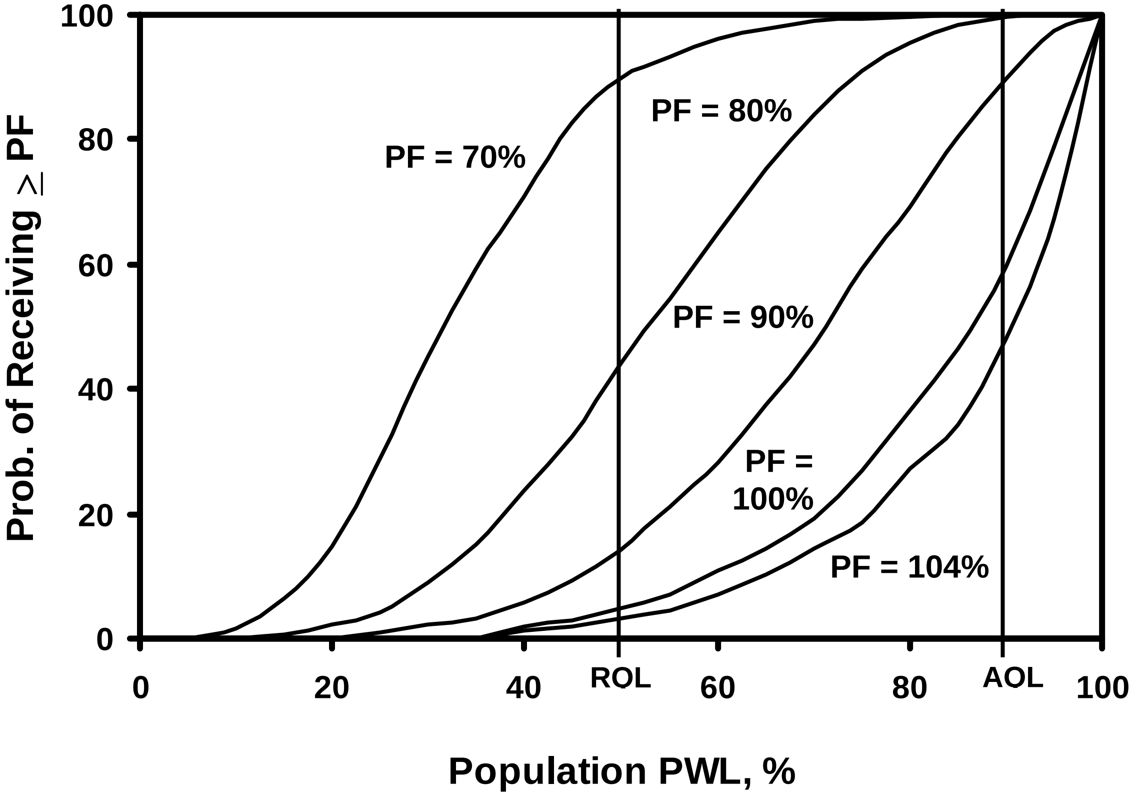


Figure 41. OC Curves for a Sample of Size 4 Using the Payment Relationship in Equation 28

Recommendation

When a price adjustment acceptance plan is used it is essential that the agency develop both an EP curve and OC curves for the probability of receiving various payment factors over the total range of quality levels. The agency may also wish to look at histograms of individual payment factors to obtain a picture of how much variability is associated with the payment factor determination. This is shown in figures 37 through 40.

The decision regarding what does or does not constitute an acceptable level of risk will to a great extent be a subjective one. There is, however, one factor that is not subjective. There is generally universal agreement that the expected payment should be 100 percent for quality that is at exactly the AQL. Although it should not be confused with the statistical risk, α, the agency may wish to consider the "average payment" risk to the contractor, if the EP is less than 100 percent at the AQL, or to the agency, if the EP is greater than 100 percent at the AQL. The EP at the RQL quality level is another point that is often specifically considered.

It must be remembered that the EP alone is not a complete measure, particularly of the likelihood that any individual lot will receive a correct payment factor. The variability of the individual payment factors about the EP curve must also be considered. Ultimately, the decision regarding what constitutes acceptable or unacceptable risks rests with the individual agency. While the determination of acceptable risks rests solely with the agency, by way of the joint industry/agency task force discussed in earlier chapters, there should be contractor input into this decision.

Word of Caution

The procedures that have been presented in the previous sections, as well as the OCPLOT program that is discussed, are primarily for the case of acceptance based on a single property. When, as will often be the case, there are multiple acceptance properties it will be necessary for the agency to develop sophisticated computer simulation methods to complete a full analysis of the risks. These analyses will be quite involved and will be dependent upon the quality characteristics chosen for acceptance and whether or not a performance model for predicting service life has been adopted by the agency. Another factor that will impact the analysis is whether a composite quality measure has been developed, or whether the individual quality measures will in some way, perhaps by adding, multiplying, or averaging, be combined into a composite payment factor. All of the possibilities cannot possibly be covered in this manual.

It is very likely that the agency will need to seek outside assistance to help in developing the simulation routines necessary to fully evaluate the risks. Universities as well as other agencies and consultants who have already developed such procedures are potential sources for this outside assistance. Once the agency has developed appropriate OC and EP curves for the acceptance plan, it should supply this information to the contractors that work in the State. Otherwise, each contractor will individually be required to seek outside help to fully understand the risks associated with the new acceptance plan, and this is not very cost effective for the contractors or ultimately for the agency.

41 Modify Acceptance Plan, if Necessary

If the risks are considered unacceptable they are likely to be too high rather than too low. To reduce the risks it may be possible to change the specification limits, the acceptance limits, and/or increase the sample size. The most straightforward approach would be to increase the sample size per lot.

An increase in the sample size may be accomplished by either increasing the lot size or increasing the sampling frequency. For example, if the lot size were 1800 Mg of HMAC, and the sampling frequency was one test per 450 Mg, then the number of tests per lot could be increased from 4 to 8 by increasing the lot size to 3600 Mg. On the other hand, the number of tests per lot could be increased from 4 to 8 by keeping the lot size as 1800 Mg but increasing the sample frequency to one test per 225 Mg.

Another way to change the risk levels would be to change the specification limits or the acceptance limits. This may be related to the definition of AQL and/or RQL material. For example, for the example presented above for asphalt content, the AQL was defined as a population with a mean of 6.00 percent and a standard deviation of 0.20 percent. Using this definition, the specification (and, in this case, acceptance) limits for an accept/reject decision based on a single test result were set at plus or minus two standard deviations from the target value of 6.00 percent, i.e., 6.00 percent ± (2 × 0.20 percent) or 5.60 percent to 6.40 percent. This provided an α risk to the seller of 0.0456, or 4.56 percent. This risk can be reduced to nearly zero by setting the specification (and acceptance) limits at $\pm 3\sigma$ rather than $\pm 2\sigma$. However, this will also increase the β risk**,** unless the definition of RQL is changed.

For accept/reject acceptance plans based on PWL, the acceptance limit could be reduced, say from 90 PWL to 85 PWL, to lower the α risk. It could also be raised, say from 90 PWL to 95 PWL, to increase the α risk. It must be noted that whenever α is changed, β will also change unless the sample size is changed as well. For acceptance plans with price adjustments, the payment equation could be changed to increase or decrease the expected payment values. While these changes would impact the EP values and the "payment" risks at the AQL and RQL, they would not necessarily change the "statistical" risks, α and β.

New OC curves and EP curves must be developed for any changes that are proposed to the initial acceptance plan provisions. This is the only way to determine what impact the changes will have on the risks to both the contractor and the agency. The agency should not proceed with developing the finalized draft specification until an acceptance plan has been developed for which the agency believes the risks are appropriate.

42

Finalize the Initial Draft Specification

Once all of the preceding steps have been completed, the agency can move forward to finalize the wording for the initial draft specification. At this point the agency is ready to move forward to the implementation phase of the specification development process.

It is obvious from the above discussions that a great deal of thought should be put into the development of an acceptance plan. There are many "pieces" to the puzzle that must fit together for the acceptance plan to be well–written and to work as intended. However, there are many resources that can be used to help accomplish this goal. QA acceptance plans have been under development and evolution for over three decades. This history can be an invaluable resource for any agency that is in the process of developing QA acceptance plans.

CHAPTER 8. IMPLEMENTATION

This chapter presents Phase III, Implementation, of the specification development process. The purpose of this chapter is to provide prescriptive steps that have been used by agencies and that may be useful as an aid in the implementation process. There are many items that can be used in implementing the specification. Some of these may be unnecessary if the specification has been used previously and only a modification is needed. On the other hand, if it is the first use of a QA specification, all of these items should be considered. The primary theme in these suggested steps is to "ease the pain" and to make the learning curve more easily understood by both the private sector and the agency personnel. The steps that are involved in the implementation process are identified in the flowchart in figure 42. The numbers in boxes before the titles of the following sections refer to the corresponding box in the flowchart.

1 Simulate Specification

Without having to do any fieldwork it is possible to use data from several projects either under construction or recently completed to investigate whether the specification works as intended. This will allow the determination of how the acceptance procedures and payment factors would have performed had the new acceptance plan been used on these projects. However, there are potential drawbacks to putting too much emphasis on these data. Caution is urged against drawing conclusions from this analysis because the contractors did not respond under these projects in the same manner as they would have had the new acceptance provisions been part of the project contracts. The data will probably not have been gathered in the same manner on the ongoing projects as will be required in the new acceptance plan. In other words, sublot and lot sizes and sampling frequency will likely be different. Nonetheless, the analysis can provide inexpensive, early, general insights as to how the new acceptance plan might perform.

Also, be sure that random sampling was used to collect the data. This analysis will only be valid if random sampling was used on the projects. Similarly, assure that the same sampling and testing procedures as decided upon in Phase II were used to collect the data. The analysis will only be valid if the same sampling and testing procedures were used that are contained in the new acceptance plan.

When analyzing the data, apply the individual and composite payment factors and critically examine the outcome. The data must be analyzed and the payment factors applied in the same manner as will apply in the new acceptance plan for the analysis to provide meaningful information concerning the new plan. Examine the payment factors in concert with the AQL and RQL and OC and EP curves to assure that incentives would have been paid for only truly exceptional quality work and that if price reductions would have occurred, that they would have been commensurate with less–than–acceptable quality.

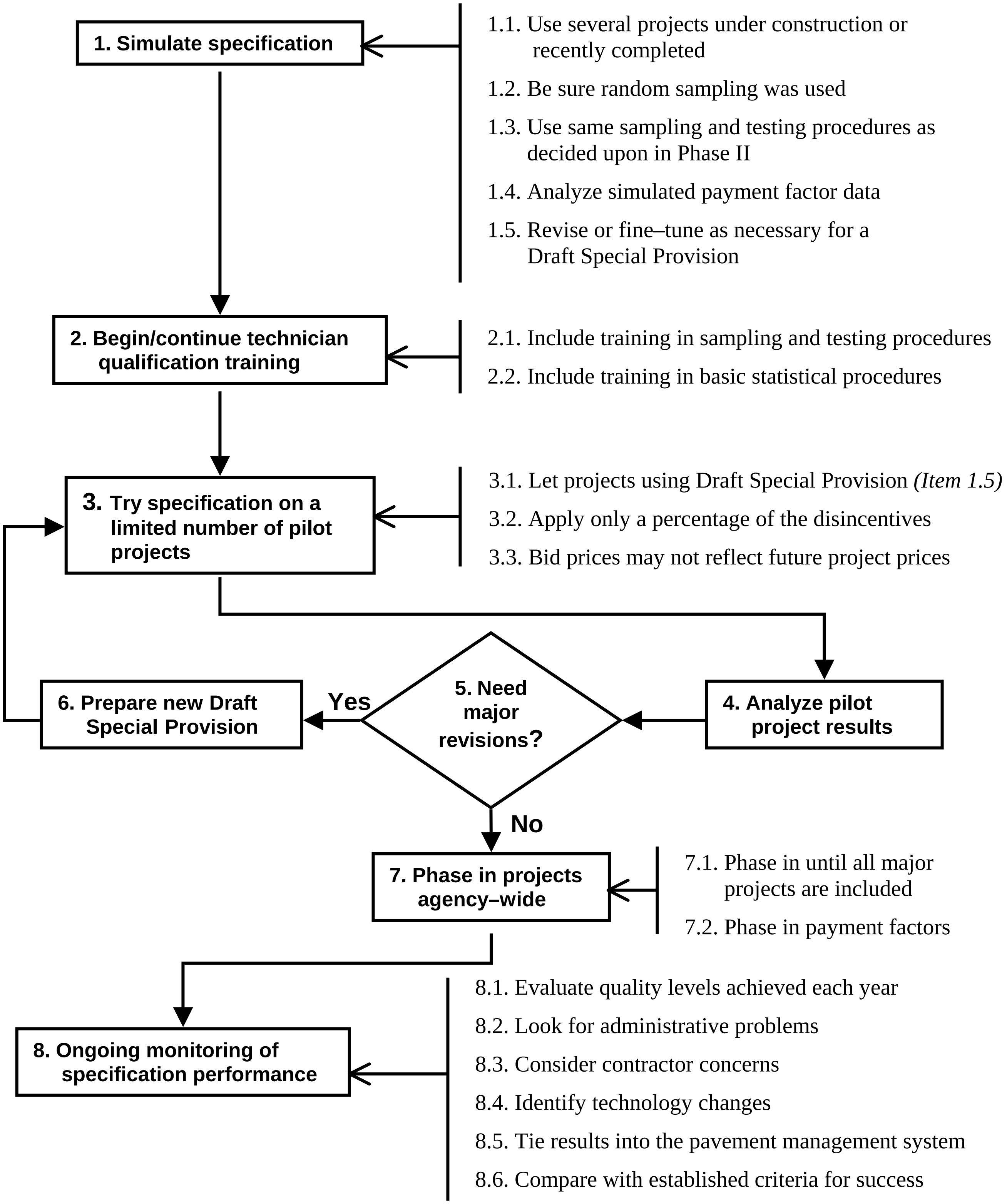


Figure 42. Flowchart for Phase III—Implementation

Another approach that may be useful is to do “shadow projects” where the project is governed by the current specifications and acceptance plans and those are used in all project decisions, but, in addition, testing is also conducted to simulate the new or proposed acceptance plans. This can be done to evaluate whether the lot size, sublot size, sample size, and other test requirements are realistic.

Once information is available from the specification simulation, revise or “fine–tune” the acceptance plan as necessary into a draft special provision. Depending on the outcome of the analysis, no changes may be necessary or it may be possible to make small changes to the specification to correct an obvious problem. However, major changes based on small problems should be avoided because the small problems may not exist when the contractor bids the new special provision and works with it as part of the contract.

2 Begin or Continue Technician Qualification Training

Technician training and qualification is an extremely important aspect of QA specification implementation. This item often has not been given the attention it needs and deserves. It is important for the technicians, both those from the contractor and those from the agency, to know how to properly perform the tests, to be completely familiar with the sampling and testing frequencies, and to have at least a general understanding of the statistical procedures that will be used to analyze the data and to perform the payment factor calculations. The training should include assurance that the sampling and testing procedures to be used are completely understood with hands–on demonstrations of competence.

Ideally, technician training and qualification will have been previously begun in preparation for this stage of the specification development process. Agencies have found that technician qualification is a necessary and time–consuming step. Such concepts as sharing responsibility between agency and contractor may be entirely new and may require time for technicians to assimilate.

Include basic statistical procedures in the training. The QA initiative will certainly contain some form of statistical procedure. More than likely this will be new to technicians. These procedures are often perceived as being too complicated for technicians to master. Training is important to dispel this notion by explaining the basic statistical procedures that are required to control and/or accept the product, and to have the technicians perform the calculations. The technicians do not necessarily have to know the statistical derivations of the new QA specification, but they should have an idea why the statistical procedures are necessary and how sampling and testing affect the outcome of the analysis.

The training is also not a onetime occurrence. As new technicians come into the employment of the contractor and the agency, training must be given to them. Also, it is generally recognized that retraining, which may require requalification, is also desirable to assure that “bad habits” do not encroach on the techniques and that technicians are familiar with new procedures as they are adopted.

3 Try Specification on a Limited Number of Pilot Projects

The next step in the implementation process is to let pilot projects using the new specification, probably as a Draft Special Provision. Letting a limited number of pilot projects of medium size allows the contractors to develop bidding strategies and to determine how the special provision works. This also allows the agency to examine the outcome and, if desirable, further "fine–tune" the specification under "real–world" conditions.

> ***Warning***
> Letting an untested special provision agency–wide without first testing it on pilot projects is not a good idea and should not be done.

As mentioned previously, the contractor is most interested in the bottom line, i.e., what is the final payment factor? To ease the uncertainty that may exist in the contractor's mind, and thus in the bidding, initially applying only a percentage of the disincentives has proven to be effective. There have been several strategies used to "ease into" a payment schedule and reduce price adjustment concerns on the part of the contractor, particularly on the pilot projects. Some agencies have applied only a portion of the incentive and disincentive. Others have found that applying only part of the disincentive but all of the incentive helps implementation. There is no consensus regarding how to best implement this strategy, but some effort to do this is recommended.

It should be realized that bid prices obtained on the pilot projects might not reflect future project prices. As with any new specification, the contractors may increase unit bid prices because of unknown risks that they feel may exist. This concern may be reduced if the task force has done a good job of communicating the specification details and processes to the industry during the development phase. An agency should not be discouraged if the initial prices are higher than normal since experience has shown that QA specifications do not increase bid prices over the long–term. Stating this differently, there are usually several other factors that affect competitive bidding to a greater extent than the use of QA specifications.

4 Analyze Pilot Project Results

All of the pilot projects should be examined in their entirety. An attempt to draw conclusions from one or two positive or negative outcomes should be avoided. The pilot projects are part of the learning process and mistakes may be made. One of the purposes of pilot projects is to restrict the mistakes to a small portion of the production/construction processes.

Some questions that should be addressed in the analysis of the pilot projects are similar to those examined under the previous specification simulation process. They include:

- Do the quality levels achieved and payment determinations appear to be consistent with those anticipated in the development of the specification?
- Do any projects stick out as being out of line with the others? For instance, did one project have much more severe price adjustments than the others? Did all projects have an incentive, irrespective of perceived quality? If any project does stand out as being exceptionally good or poor, the reason should be sought.
- Did contractors have trouble meeting the special provisions due to geographical location? As an example, are the available materials conducive to meeting the requirements?
- Did contractors have trouble meeting the special provisions due to technological reasons?
- Does the agency view the specifications as being too liberal or too severe in relation to the quality provided?

5

Do the Draft Special Provisions Need Major Revisions?

Do any of the problems that were found in the analysis of the pilot projects indicate that major revisions should be made to the draft special provisions?

If the identified problems are substantial and minor revisions will not correct them, then a new draft special provision should be considered. If the steps along the way have been taken with care and forethought, the problems should not be major, but it can happen. If no major problems are found, the next implementation step is to proceed to phasing–in the specification on projects agency–wide.

6

Prepare New Draft Special Provisions if Necessary

If some unfortunate occurrence requires new draft special provisions, the same steps taken in chapters 3 through 7 must be examined to see how the problem can be solved. At the very least, the timeframe for implementation will be adversely affected. However, not correcting the flaw will most likely negatively influence the specification to a greater extent and for a longer period of time.

7 Phase In Specification on Projects Agency–Wide

There is no single scenario that is applicable to all agencies. One note of caution is not to rush the process. It takes time for contractor and agency personnel to become familiar with and comfortable with QA specifications.

Phase in the specification until all major projects are included. Starting with a limited number of medium–sized projects and proceeding to all projects over 2 to 3 years allows contractor and agency personnel to become comfortable with the specification before having to bid large projects.

Consider phasing in payment factors. Phasing in payment factors over the 2–year to 3–year period discussed above in regard to the pilot projects may aid in the implementation process. A schedule of 50 percent of the intended payment adjustment the first year and 100 percent the second year, or 33 percent, 67 percent and 100 percent over 3 years, are two possible ways of phasing in the payment factors. These phased in payment factors may apply to both incentives and disincentives, or to only the disincentives.

8 Monitoring Program for Specification Performance

An annual review or monitoring of the specification is desirable. Some of the items to be examined in the monitoring program include:

- **Evaluate quality levels achieved each year.** For each quality characteristic, project data should be analyzed to determine how well the process standard deviations that are being attained compare with the "typical" values used when developing the specification. The project standard deviation values should be computed in a manner consistent with the way in which the typical standard deviation values were determined. Inconsistencies, either too high or too low, between the standard deviations being obtained and those used to develop the specifications, may indicate that revisions to the specifications are necessary.

  Project data should also be analyzed to determine whether or not the populations being obtained on projects are meeting, failing to meet, or exceeding the AQL definitions on which the specifications were based. Data should also be gathered and analyzed concerning payment factors that are being awarded on projects. The payment factor data and the project population data should be compared to see how they relate to the OC and EP curves that were developed for the specifications, and also to identify if there are inconsistencies between the quality that is being achieved and the payment factors that are being awarded.

If the desired quality, i.e., the AQL, is not being achieved this may indicate that the technology is not sufficient to allow the contractors to achieve the AQL. However, it may also indicate that it is more economical for the contractor to operate at a quality level less than the AQL. This could be because full payment can be obtained with less than AQL quality, or it may be that the payment schedule is such that it is cheaper for the contractor to operate below the AQL and accept the corresponding payment disincentive than it is to operate at the AQL. In either case, modifications to the payment schedule might need to be considered.

Ideally, the quality levels will gradually improve or remain stable with time. If the quality levels start a downward trend or increase significantly, a thorough analysis of the reasons for this is imperative. While the significance might not be as obvious as for a decrease in quality levels, an increase in quality levels is also an indication that the specification may need to be modified. It may be that technology changes are responsible for the improved quality, and the specifications should be revised to reflect this situation.

- **Look for administrative problems.** Is the definition for a lot creating any problems such as having lots open for long periods before the payment factor can be determined? Have the acceptance procedures created excessive paperwork? Is the agency having trouble applying incentives or disincentives? Are there other administrative problems? If so, these should be identified and corrected.

- **Consider contractor concerns.** Does the industry think the specification is working as intended? Do they have constructive suggestions as to how it can be improved? Are there problems finding enough qualified technicians? Listen to the industry's concerns. If these concerns are not heard and addressed, the industry may bring pressure to bear that may undo the advances made with the new QA specification.

- **Identify technology changes.** As stated previously, the specifications should be dynamic. Have technology changes occurred that can positively impact the specifications?

- **Tie results of the QA specification into the pavement management system.** For the specification to be a true measure of quality, closing the loop of relating quality and performance requires the integration of the quality levels and/or payment factors with the agency's pavement management system. This may be a source of performance information for future modifications of the specifications.

- **Compare the quality levels with established criteria for success.** Close the loop with the item in Phase I in which goals and expectations were established. Is the specification performing the way it was anticipated? If it is not, why not? Are further revisions needed?

CHAPTER 9. CASE STUDIES

This chapter presents some case studies of the development, implementation, and evolution of QA specifications by the NJDOT. New Jersey was one of the first states to investigate and implement statistically based, QA specifications. As such, the NJDOT has a long history with QA specifications. These case studies have been selected primarily because the authors are very familiar with all of the steps involved and all of the specific details of the processes used in developing these specifications.

Background to the Case Studies

These case studies are intended to provide examples of how some QA specifications were initially developed and how they have evolved through the years. They are presented only to show the steps that are involved, the thought process that might be followed, and the types of decisions that must be made when developing QA specifications. They are NOT intended to represent the only, or even the best way, for any individual agency to develop its own QA specifications. These case studies have been selected for presentation because they clearly illustrate many aspects of the acceptance plan development process that are presented and discussed in the previous chapters and in the appendices.

Development of Statistical Expertise

As noted in the flowchart in figure 4, step 7.6, a decision that must be made very early in the specification development process is whether or not there is sufficient expertise within the agency, or if it will be necessary to seek outside assistance. This usually poses no problem as far as design, construction, inspection, and testing are concerned, because transportation agencies usually have many employees who are well trained in these areas. A critical area of expertise that often is not present, however, is statistical engineering.

Because this was recognized to be a vital prerequisite for a sound QA program, the NJDOT decided at the outset that it would be preferable to have the necessary statistical expertise as part of its in–house staff. Consequently, a new set of job specifications—the Statistical Engineer Series—was created to fill this perceived gap. Table 23 outlines the educational and experience requirements for the entry, intermediate, and supervisor levels; and the complete job specifications are contained in appendix N.

Table 23. Requirements for NJDOT Statistical Engineer Series

	Entry Level	Intermediate Level	Supervisor
Engineering Degree	BSCE or equivalent	BSCE or equivalent	BSCE or equivalent
Applied Statistics Credits	12	18	24
Computer Science Credits	6	6	6
Experience, years	2	3	4

Basic Philosophy—Simple But Scientific

The overriding philosophy of the NJDOT QA program is to use methods that are "simple but scientific," leading to the development of statistical construction specifications that

- Are scientifically and mathematically sound.
- Are related to performance.
- Are easy to understand and apply.
- Provide strong incentives to produce good quality.
- Provide strong disincentives for poor quality.

The first objective has been accomplished by using well–established statistical methods, and performing OC curve and EP curve analyses to verify that the acceptance procedures perform as intended. The second objective has been met by using pavement design methodology, engineering judgment, or a combination of both to develop prototype mathematical models to predict expected life from as–constructed quality levels. The third objective has been accomplished by choosing the simplest approaches possible and switching to more complex methods only if real data indicate the simple methods are inadequate in some way. The fourth objective has been achieved by recognizing both the technical and psychological benefits of positive–incentive provisions that reward excellent performers with bonuses in addition to the contract bid price. And, finally, the fifth objective has been accomplished by using life–cycle cost analysis methods to justify sufficiently severe payment reductions to strongly discourage poor quality workmanship.

Implementation Strategies

A factor that must not be overlooked is that the construction industry is an indispensable partner in any highway agency's QA program. For any quality program to be successful for any extended period of time, it is essential that it be understood and supported by the contractors, producers, and suppliers who must work with it on a daily basis. This means that it must make sense to knowledgeable people in the field and must be perceived as fair and effective.

As this manual has previously emphasized, the surest way to guarantee that these conditions will be met is to include all the stakeholders in the specification development process. An approach that has worked extremely well for the NJDOT has been to form joint task forces that include representatives from the highway agency, FHWA, and the construction community. The specification and acceptance procedure should be well thought–out and thoroughly analyzed (OC curves, EP curves, etc.) before it is presented to the task force for their review since technical details are not easily worked out by group discussion. The agency can greatly enhance its credibility by being open to constructive criticism, and by providing valid, supportable answers to the many questions that will arise. By acknowledging that the specification is a prototype, and making the commitment to carefully monitor field performance and consider any necessary changes, the agency can expect to receive the support of the construction industry to proceed with a series of pilot projects. An additional condition that may be desirable for the field trials is to scale back the payment–adjustment provisions for the pilot projects, perhaps by as much as 50 percent. This allows both agency and contractor personnel to become more familiar and comfortable with the specification before it goes fully into effect.

Fundamental Principles

The following are some of the basic principles of statistical specification writing, listed in the approximate chronological order of their discovery:

- Recognition that highway construction is highly variable.
- Recognition that PD (or its complement, PWL) is a better measure of construction quality than the average value by itself.
- Understanding that construction of the OC curves, or the EP curve, is the only way to be sure an acceptance procedure is performing as intended.
- Recognition that the standard deviation is a better measure of variability than the range.
- Recognition that some degree of bonus provision is required in order for specifications based on PD (or PWL) to properly pay an average of 100 percent at the AQL.
- Recognition that continuous (equation–type) payment schedules tend to avoid disputes over test precision or round off rules, etc.
- Development of the life–cycle cost basis for payment schedules.
- Development of composite payment equations to provide a practical method to determine the net payment factor for multiple quality characteristics.
- Application of straightforward methods to develop approximate mathematical models to predict expected life for performance–related specifications.
- Development of the composite quality measure to provide a still more convenient method for dealing with multiple quality characteristics.

These principles, along with basic sampling and testing methodology, are applied in the case studies that follow.

Choice of Statistical Quality Measure

Like most agencies, the NJDOT recognized that both the average level and the variability of construction work will affect expected performance. AAD and CI were not seriously considered as potential statistical quality measures, so this narrowed the range of possible choices to PD and PWL. Both are exactly equivalent from a mathematical standpoint, one being the complement of the other, but the NJDOT chose to use PD for specific reasons.

The primary reason was that it was the measure used in the original source documents. (23) A second reason was that the computation of total PD for a two–sided specification is slightly simpler and more intuitive than it is for PWL. And, finally, the use of PD casts the payment equation in a form that is easier to interpret by inspection, i.e., the constant term represents the maximum payment factor at PD = 0, and payment factors for increasing levels of PD gradually decrease from that maximum.

Case Study 1: Portland Cement Concrete (PCC) Specification

The PCC specification was the first of the modern statistical specifications developed by the NJDOT. What "modern" means is that most or all of the fundamental principles of statistical specification writing listed in the previous section are applied in a scientific manner. (The last three of this list were not felt to be necessary for the PCC specification but are included in case study 2.)

Acceptance Quality Characteristics

PCC acceptance is based on three quality characteristics—slump, air entrainment, and compressive strength. Since slump and air entrainment can be measured at the jobsite before the concrete is placed in the forms, the decision was made to use a pass/fail procedure for these measures. It was further decided that, if either slump or air entrainment (or both) were below the desired level based on the first set of tests, a single retempering would be permitted to attempt to bring the mix into the acceptable range. Typical acceptable ranges were ±25 mm for slump and ±1.5 percent for air entrainment, both of which represent ranges of about plus or minus two standard deviations.

Since compressive strength could only be evaluated after the concrete had cured, this characteristic was well–suited for acceptance by payment adjustment. Design strengths for four different classes of concrete are listed in table 24. The AQL was typically defined as PD = 10 percent below the class design strength (equivalent to PWL = 90).

Payment Considerations

It was the initial development of this specification in the late 1970s that led to a request to the Attorney General's office to furnish an opinion concerning the legality in the State of New

Jersey of paying bonuses for superior quality. A review of existing statutes turned up nothing that specifically prohibited such an application so that, in the opinion of the Attorney General, such provisions were permissible. It was cautioned that it would be advisable to have evidence that extra quality, beyond what was specified in the contract documents, resulted in extra value to the State and the motoring public, thus justifying the use of bonus clauses. (This provided additional impetus to efforts already under way to establish quantified relationships that would ultimately be useful in developing performance–related specifications.)

Upon receipt of this opinion, it was decided to use a relatively modest bonus provision that provided a maximum payment of 102 percent for PCC items that tested out at the highest possible quality level of zero PD. Since the payment factors were applied to the in–place cost of the concrete, and concrete items tend to be relatively expensive, this provided substantial bonuses in many cases.

Table 24. Classes of Concrete for NJDOT Specification

Class	**Class Design Strength**		**Typical Use**	**Pay Adj. Item**	**Tests per Lot** [1]	
	PSI	**MPa**			**Initial**	**Retest** [2]
P [3]	5500	37.9	Prestressed members	Yes	5	5
A	4200	29.0	Bridge decks, columns, etc.	Yes	5	5
B [4]	3700	25.5	Pavement, footings, etc	If Yes: If No:	5 2	5 5
S	2000	13.8	Seal (tremie)	No	1	—

[1] A test is defined as the average strength of a pair of compression cylinders.
[2] A retest, when required, is defined as the strength of an individual core.
[3] Higher levels of Class P are frequently included by special provision.
[4] Some Class B items may be declared as payment adjustment items.

More recently, the decision was made to increase both the positive incentives and the negative disincentives of the PCC payment schedule. NJDOT specification writers prefer to express payment schedules in terms of payment adjustment rather than payment factor and, accordingly, equations 29 and 30 were developed. As can be seen from equation 29, the AQL of 10 PD corresponds to a payment adjustment of zero (100 percent payment). The maximum bonus payment provided by equation 29 is +3.0 percent, while the minimum payment adjustment assigned by equation 30 is –50 percent.

$$\text{if } PD < 50: \quad PA = 3.0 - 0.3\text{PD} \qquad (29)$$

$$\text{if } PD \geq 50: \quad PA = 26.0 - 0.76\text{PD} \qquad (30)$$

where: PA = payment adjustment (percent).
PD = lot percent defective.

The payment schedule given by equations 29 and 30 illustrates a feature frequently used by the NJDOT, which has been well received by the construction industry. Provided that the quality level does not deviate too far from the desired level, the effect on the amount of payment adjustment is relatively moderate (equation 29). However, for seriously deficient quality, the level of payment falls off much more rapidly (equation 30).

Like many other agencies, the NJDOT also defines an RQL. This provision supersedes the payment equation and allows the agency the option to require removal and replacement whenever a poor quality level might severely impair the performance of an item. Based on the fatigue relationship of the AASHTO design equation, and consultations with pavement engineers, it is believed that a percent defective level of PD = 75 might result in a loss of service life of about 50 percent for New Jersey's PCC pavement design.(31) Based on this assumption, the RQL was set at PD = 75 with the provision that, if the agency elects not to enforce it, then the payment schedule given by equation 30 applies.

Ideally, it would be desirable to also have fatigue relationships for bridge decks and other structural items, similar to the AASHTO design equation for pavement, making it possible to predict the service lives of these items as a function of the as–constructed quality. An ongoing research study will attempt to determine if sufficient performance data are available to establish reliable relationships of this type. In the meantime, until such relationships can be developed, the NJDOT has decided to use the payment equations developed for pavement and apply them to all concrete items.

Unlike the application of payment adjustments, for which underestimates and overestimates of quality tend to balance out in the long run, there is no such compensating property associated with estimates of RQL material. In other words, it would not be considered an appropriate result if an agency falsely rejected a lot or two that were not actually as poor as RQL quality and, at the same time, mistakenly accepted a lot or two that truly were as poor as RQL quality. It is desired to keep the risk of either type of mistake at manageably low levels, and the easiest way to do this is to employ a retest to confirm the result when a seriously out–of–specification condition is detected. In the case of the NJDOT specification, a retest option occurs at the breakpoint of the payment schedule at PD = 50. Retest sampling rates are listed in table 24.

As noted in table 24, not all concrete items are accepted by payment adjustment. Some less critical items, usually Class B concrete, may be accepted by a pass/fail procedure. Provided that no individual strength test result falls below an appropriate limit, the item is accepted at full payment. Although it does not occur often, when a non–payment–adjustment item is rejected by the pass/fail procedure, it is then subject to retest and, from that point on, is treated as a payment–adjustment item.

The earlier version of this specification was implemented in the late 1970s, first on a series of pilot projects and, when that proved successful, it was included on all jobs thereafter. Initially, the construction industry was very apprehensive about the specification. Bid prices were somewhat higher than normal on the first few jobs, and some contractors hired testing laboratories to take companion cylinders to compare to the acceptance cylinders taken by the NJDOT. Also, it was obvious from the unusually high strengths obtained on the pilot projects (increase of about 7 Mpa) that contractors were being quite conservative with their mix designs.

As more experience was gained on the pilot projects, it was apparent that the independent testing labs were obtaining strengths that compared favorably with NJDOT results, and most contractors discontinued taking companion cylinders. Also, because of the conservative mix designs, strengths were higher than necessary and most contractors were earning bonuses. As time went on, the construction industry realized that with moderate care it could easily meet the requirements of the specification and bid prices tended to move back toward normal levels. Mix designs strengths tended to stay higher than in the past, but not as high as they were on the first few pilot projects.

When the latest version of the specification was implemented, the concern on the part of the construction industry was considerably less than it had been for the initial version. Still, it was deemed advisable to first implement the specification on a series of pilot projects. Once again, the experience was very favorable with the majority of lots earning full bonus, and the NJDOT plans to implement the specification on all future projects.

Case Study 2: Superpave® HMAC Specification

Background

The development of the NJDOT Superpave specification consisted of three major phases. For the first phase in the early 1990s, the Superpave design method had not yet been developed. (24) At that time, the NJDOT was controlling the in–place density of its HMAC with an acceptance procedure based on the average air voids level determined from a series of cores taken from the completed mat. Although this specification had worked reasonably well for a number of years, it was noticed that the average often fell on the high side of the allowable range of 2.0–8.0 percent, and that individual values falling well above the upper limit of 8.0 percent were quite common. Consequently, it was decided to revise the specification and use a different quality measure, PD, designed to simultaneously control both the average level and the variability of the items to which it is applied.

Changing from an acceptance procedure based on the mean to one based on PD essentially redefined the AQL from a very lenient value of PD = 50 to a considerably more demanding level of PD = 10. Other modifications at the time included switching from a stepped to an equation–type payment schedule, adding a bonus provision for superior work, and including a remove–and–replace clause for seriously defective work.

A procedure and payment schedule, very similar to that described for the PCC specification in case study 1, were developed. Several pilot projects were constructed, quality levels were consistently good, and the specification was then broadly implemented.

In the late 1990s, a second phase of pilot projects introduced the change to the Superpave design method and, at the same time, the incorporation of a composite payment schedule based on three quality characteristics—in–place air voids, thickness, and smoothness. (25) Individual compound payment schedules were developed for each of these three characteristics, all based on PD as the statistical quality measure. Each of the payment schedules awarded bonuses for excellent quality, assessed moderate reductions for quality that strayed minimally from the desired level, and imposed more severe reductions for seriously deficient quality. Retest and remove–and–replace provisions were also included in this version of the specification. The composite payment equation combined the results of the individual payment equations, and the overall lot payment factor was taken as the average of the individual payment factors.

Several pilot projects were constructed with the phase two version of the specification. Although the quality ranged from good to excellent, with very few poor quality lots, it soon became apparent that certain features of the acceptance procedure were not optimal and should be improved. Thus began the third phase of the development of the Superpave specification.

Payment Considerations

A primary concern was that the method of averaging the payment factors for the individual quality characteristics did not accurately reflect the effect of the individual quality measures on pavement performance, thus falling short of relating the specification to performance. If the true performance model for HMAC pavement as a function of the three acceptance characteristics were known, it is believed that it would behave approximately as indicated in table 25.

Table 25. Intuitive Model for Air Voids, Thickness, and Smoothness of HMAC Pavement

Quality Levels	Service Life
Poor in any one characteristic	Some loss
Poor in any two characteristics	Greater loss
Poor in all three characteristics	Substantial loss

If the intuitive model in table 25 is even approximately correct, the method of averaging the individual payment factors in the prototype specification will not withhold sufficient payment to cover the likely costs of future repairs, nor will it pay an appropriate amount of bonus for truly superior quality. If the specification is to be truly related to performance, an additive process must determine the lot payment factor, although not necessarily a direct or fully additive process.

To confirm and approximately quantify the intuitive model in table 25, the NJDOT conducted a national survey to obtain estimates of pavement life for selected combinations of quality levels for the three measures—air voids, thickness, and smoothness. Responses were received from 35 States, of which four indicated that this information was not available, while another five failed a consistency check. This left a total of 26 responses, the averages of which are presented in table 26.

Table 26 provides the data from which it is possible to infer how experienced engineers believe various combinations of poor quality in the three characteristics affect pavement service life. The first row represents the reference condition, an expected life of 20 years when all quality measures are at their respective “good” levels. The next three rows include those cases for which just one of the three measures is at the “poor” level, producing an average expected life of (16.1 + 15.0 + 11.6) / 3 = 14.2 years. The next group of three rows includes those cases for which two of the three measures are at the “poor” level, producing a lower average expected life of (11.9 + 9.3 + 8.7)/3 = 10.0 years. The final row in table 26 represents the case for which all three measures are at the “poor” level, and produces a still lower expected life of 6.8 years.

Table 26. Pavement Performance Survey Results

Initial Quality Levels			Expected Life (Years)
Smoothness	**Air Voids**	**Thickness**	
Good	Good	Good	20.0 [1]
Good	Good	Poor	15.0
Good	Poor	Good	11.6
Good	Poor	Poor	8.7
Poor	Good	Good	16.1
Poor	Good	Poor	11.9
Poor	Poor	Good	9.3
Poor	Poor	Poor	6.8

[1] Given as a reference point.

Since these results were consistent with the intuitive model in table 24, it was decided to proceed with the development of an additive model for this specification. Since it was planned to change to a new type of smoothness measure in the near future, it was decided to use only air voids and thickness in the composite payment schedule to be developed, and to treat smoothness separately.

At the time this specification was developed, the exponential model described in chapter 6 had not yet been developed, so the polynomial model for two characteristics given by equation 31 was used.

$$EXPLIF = C_0 + C_1 \times \mathrm{PD}_{VOIDS} + C_2 \times \mathrm{PD}_{THICK} + C_3 \times \mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK} \tag{31}$$

where:
$EXPLIF$ = expected life (years).
PD_{VOIDS} = air voids percent defective.
PD_{THICK} = thickness percent defective.
C_i terms = coefficients to be determined.

To determine the four unknown coefficients for this model, it is necessary to have four separate pieces of performance data, as shown in table 27. These values represent a consensus based on NJDOT field experience, plus an analysis using the AASHTO design procedure for flexible pavement using typical NJDOT data. (31)

Table 27. Matrix of Expected Pavement Lives Used to Develop Polynomial Performance Model

	Thickness Quality	
Air Voids Quality	PD = 10	PD = 90
PD = 10	20 yrs.	10 yrs.
PD = 75	10 yrs.	5 yrs.

The four pieces of data in table 27 were then used to write four simultaneous equations that were solved to obtain the four unknown equation coefficients, producing the performance model given by equation 32.

$$EXPLIF = 22.9 - 0.163\ \mathrm{PD}_{VOIDS} - 0.135\ \mathrm{PD}_{THICK} + 0.000961\ \mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK} \tag{32}$$

There are two ways that this model can be used to develop an acceptance procedure and adjusted payment equation. It can be used directly, first to estimate expected life, which is then substituted into a payment equation to determine the payment adjustment. Alternatively, it can be converted into a composite quality measure (see complete development in appendix K), in which case the payment adjustment is expressed as a function of the composite quality measure. For this specification, the NJDOT chose the latter approach, producing the expression for composite quality measure (PD*) given by equation 33:

$$\mathrm{PD}^* = 0.807\ \mathrm{PD}_{VOIDS} + 0.669\ \mathrm{PD}_{THICK} - 0.00476\ \mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK} \tag{33}$$

The use of the composite quality measure also provided an effective way to resolve another serious concern with the existing specification—the fact that the RQL provision was inconsistent in the way it dealt with simultaneous failures in two or more acceptance characteristics. The purpose of the RQL provision is to allow the agency the option to require removal and replacement, at the contractor's expense, of an item whose quality is so deficient that its performance may be severely impaired. Initially, the RQL for both air voids and thickness had been defined as PD ≥ 75 so that if either air voids or thickness exhibited this level of quality the lot was considered rejectable. Later, based on attempts to relate quality to performance, it was decided that thickness PD could be as large as PD = 90 before such a severe consequence was justified. However, a more troubling problem was the inconsistency shown in table 28.

Table 28. Inconsistency in RQL Provision in Prototype Specification

Case	**Quality Level**		**Rejectable ?**
	Air Voids	**Thickness**	
1	PD = 10 (AQL)	PD = 90 (RQL)	Yes
2	PD = 75 (RQL)	PD = 10 (AQL)	Yes
3	PD = 74 (Almost RQL)	PD = 89 (Almost RQL)	No

In cases 1 and 2, if either air voids or thickness is at the RQL while the other is at the AQL of PD = 10, this produces a reject condition. Case 3, in which both characteristics are almost at the RQL, clearly is far worse than the other two cases but does not lead to a reject condition.

Defining the RQL in a more appropriate way using the composite quality measure can rectify this inconsistency. Substituting either $PD_{VOIDS} = 10$, $PD_{THICK} = 90$ or $PD_{VOIDS} = 75$, $PD_{THICK} = 10$ into the composite quality measure equation produces $PD^* \approx 64$, which, for practical purposes, is rounded to $PD^* \geq 65$ as the RQL limit. Figure 43 illustrates how the RQL provision stated in this manner applies for all possible combinations of PD_{VOIDS} and PD_{THICK}. It is the negative cross–product term that produces the concave downward shape, and any combination of PD_{VOIDS} and PD_{THICK} falling on or above the line is considered rejectable. Note that the problematical case 3 in table 27 is clearly rejectable by this method and, also, that the curve passes through the two primary RQL points: $PD_{VOIDS} = 10$, $PD_{THICK} = 90$ and $PD_{VOIDS} = 75$, $PD_{THICK} = 10$. (The NJDOT has estimated that pavements falling anywhere on the RQL line in figure 43 have approximately a 50 percent loss of expected service life.)

The composite quality measure applies to both base and surface courses. In the NJDOT specification, total thickness is measured in conjunction with the surface layer. For base course, or for surface course of non–uniform thickness for which no thickness requirement applies, a default value of $PD_{THICK} = 10$ is used to compute the composite quality measure.

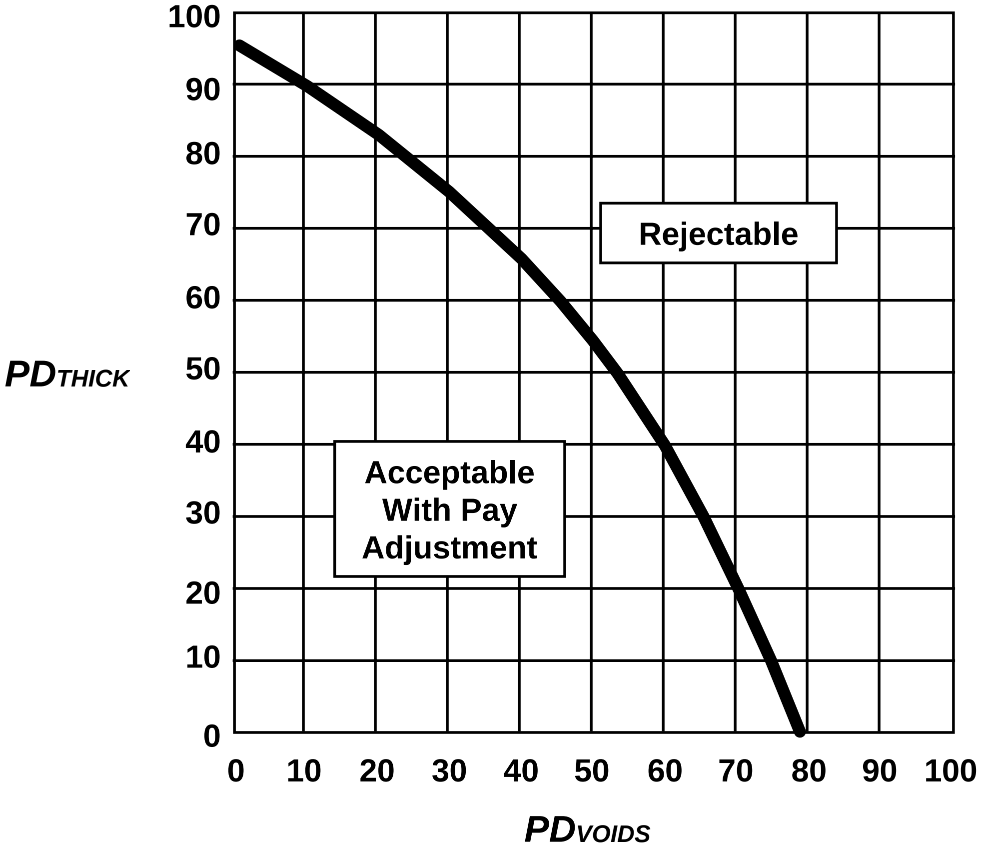


Figure 43. Graph of Composite RQL Provision Given in Equation 33

Because the expected service life of the pavement is approximately the same for any given value of PD*, this provides a convenient basis for a payment equation related to performance. Equations 34 and 35 are the equations for percent payment adjustment (*PPA*) developed by the NJDOT for air voids and thickness of mainline paving, ramps, and new shoulders in the latest Superpave specification. For existing shoulders, which experience has shown are more difficult to compact, the *PPA* computed with equation 34 or 35 is multiplied by a factor of 0.5. The lot payment adjustment is the *PPA*, expressed as a decimal, multiplied by the in–place dollar value of the lot. Alternatively, construction personnel sometimes use the *PPA*s to adjust the tonnage before totaling it up for payment, which produces the same result.

$$\text{PD*} < 40: \quad PPA = 10 - 0.67\text{PD*} \tag{34}$$

$$\text{PD*} \geq 40: \quad PPA = 116 - 3.32\ \text{PD*} \quad (\text{Minimum} = -100) \tag{35}$$

$$(\text{RETEST:} \quad \text{PD*} \geq 40, \quad \text{REJECT:} \quad \text{PD*} \geq 65)$$

These payment equations incorporate the same feature described in case study 1. As long as the contractor can maintain reasonably good control so that PD* < 40, the consequences in terms of payment adjustment are relatively moderate. If, however, the process goes sufficiently out of control that values of PD* ≥ 40 are obtained, the payment equation is much more steeply inclined and the consequences become more severe. Although it could be put at any point the agency considers appropriate, the NJDOT has found it convenient to put the retest option at the breakpoint of PD* = 40 in this compound payment equation.

Another drawback of the prototype specification was that, except for new construction, it did not include an incentive/disincentive provision for riding quality. In the near future, the NJDOT plans to switch to a profilometer–type smoothness measuring device, but, until that change is made, an interim incentive/disincentive plan has been developed for typical resurfacing projects that consist of milling followed by two paving layers. The plan is based on percent defective length (PDL), defined as the percent of the length of the pavement lot having deviations exceeding 3.175 mm in 3.048 m as measured with a rolling straightedge.

An agreement was reached with industry associations to use an incentive/disincentive plan that is less severe than is likely to be proposed when a new procedure is developed around a more sophisticated smoothness–measuring device. The interim payment schedule, based on the rolling straightedge, is given by equations 36 and 37. Pay adjustments (PA) for smoothness are calculated in units of $\$/m^2$, and the net payment adjustment for the lot is determined by multiplying by the lot area in square meters. As was done for air voids and thickness, the payment schedule consists of a compound payment equation that is moderate for quality that is reasonably under control, and more severe for quality that is clearly out of control. The payment adjustment computed for smoothness is added to any adjustment computed from the composite measure for air voids and thickness.

$$\mathrm{PD}_{SMOOTH} < 2.0: \quad PA = 0.34 - 0.26\,\mathrm{PD}_{SMOOTH} \quad (\$/\mathrm{m}^2) \tag{36}$$

$$\mathrm{PD}_{SMOOTH} \geq 2.0: \quad PA = 0.72 - 0.45\,\mathrm{PD}_{SMOOTH} \quad (\$/\mathrm{m}^2) \tag{37}$$

$$(\text{RETEST:} \quad \mathrm{PD}_{SMOOTH} \geq 2.0, \quad \text{REJECT:} \quad \mathrm{PD}_{SMOOTH} \geq 3.5)$$

Typical Results

There occasionally are resurfacing projects that do not meet the necessary criteria for thickness and smoothness to be accepted by payment adjustment. For these cases, table 29 illustrates how the new procedure applies across the full range of quality from the best (PD_{VOIDS} = 0) to the worst (PD_{VOIDS} = 100). Table 30 illustrates the performance over the same range when all three payment schedules apply.

Table 29. Performance of New Specification without Thickness and Smoothness Payment Schedules

Base and Surface		Avg. PPA (%)	Lot Payment	
PD_{VOIDS}	PD*		$/m^2	$/Lane–Km
0.0	6.7	+5.51	+0.52	+1,902
10.0	14.3	+0.42	+0.04	+146
43.9	40.0	–16.80	–1.58	–5,780
76.8	65.0	–99.80	–9.40	–34,385
100.0	82.6	–100.00	–9.42	–34,458

Note: Costs based on 5.1 cm base, 5.1 cm surface, $38.59/Mg

Table 30. Performance of New Specification When All Payment Schedules Apply

Base Course		Surface Course				V & T PPA (%)	Total Lot Payment	
PD_V	PD*	PD_V	PD_T	PD*	PD_S		$/m^2	$/La–Km
0.0	6.7	0.0	0.0	0.0	0.5	+7.76	+0.94	+3,439
10.0	14.3	10.0	10.0	14.3	0.5	+0.42	+0.25	+914
43.9	40.0	0.0	59.8	40.0	0.5	–16.80	–1.37	–5,011
43.9	40.0	30.0	30.0	40.0	2.0	–16.80	–1.76	–6,438
43.9	40.0	49.6	0.0	40.0	3.5	–16.80	–2.44	–8,926
76.8	65.0	0.0	97.2	65.0	0.5	–99.80	–9.19	–33,617
76.8	65.0	53.2	53.2	65.0	2.0	–99.80	–9.58	–35,044
76.8	65.0	80.6	0.0	65.0	3.5	–99.80	–10.26	–37,531
100.0	82.6	100.0	100.0	100.0	3.5	–100.00	–10.28	–37,604

Note: Costs based on 5.1 cm base, 5.1 cm surface, $38.59/Mg
V = Voids, *T* = Thickness, *S* = Smoothness, La–Km = Lane–Kilometer

Because table 30 includes all three measures of quality, both the incentives and the payment reductions cover a wider range than the corresponding values in table 29. In table 29, it is seen that excellent quality with zero PD_{VOIDS} produces a PD* value of 6.7 and a bonus payment of $1902 per lane–kilometer. In table 30, this same level of quality combined with excellent quality in thickness produces a PD* value of zero, which, when combined with the excellent smoothness

PD value of 0.5, corresponds to a bonus payment of $3439 per lane–kilometer. For extremely poor quality, both tables show that the maximum payment reduction is in the range of –$30,000 to –$40,000 per lane–kilometer.

The intermediate values in both tables present additional examples between these two extremes, and table 30 illustrates how different combinations of air voids and thickness PD can produce identical values of PD* (40.0 and 65.0), thus representing approximately equivalent performance from the standpoint of these two measures. It is this feature that makes the composite quality measure particularly well suited for relating performance to the payment schedules.

Remaining Issue To Be Resolved

The procedure described thus far is practical and effective, but it contains a flaw that eventually should be addressed. Like the problem with the individual RQL provisions for air voids and thickness outlined in table 26 that was resolved by defining the composite quality measure, PD*, a similar inconsistency exists with PD* and PD_{SMOOTH}. One way to resolve this problem would be to define a more complete composite quality measure that incorporates all three individual measures—air voids, thickness, and smoothness—and the general development of models for three or more quality characteristics is described in appendix L.

The NJDOT has chosen not to take this approach, however, because a major change in the current riding quality specification is imminent. When a new smoothness measuring device is selected, it is believed that it will be capable of being driven at or near traffic speed. In that case, it is likely that smoothness acceptance lots will no longer be treated in a piecemeal fashion along with air voids and thickness lots but, instead, may be defined as the entire project in one direction.

Summary of Acceptance Procedure

The new specification has payment schedules that are linked to expected performance, is easy to understand and apply, and provides increased incentive to provide good quality and avoid poor quality. To the extent possible it has been related to pavement performance through an analysis of expected service life and life–cycle costs. The ease of application can be seen in the outline in figure 44 that summarizes the key features of the complete acceptance procedure. With the use of the composite quality measure (PD*) and other refinements, the new specification requires a total of only five payment or payment–related equations to state the requirements for air voids, thickness, and smoothness. (The specification it replaced required more than a dozen payment equations to accomplish the same purpose.) And, finally, because the composite quality measure reflects the approximate additive nature of the effects of the three acceptance characteristics, both the incentives for excellent quality and the payment reductions for poor quality have been substantially increased, thus providing a strong incentive to produce good initial quality.

Another potential benefit of the increased incentive provision could foreshadow a profound change in the bidding process. Some NJDOT contractors have indicated that they have sufficient confidence in their QC operations and their ability to earn the incentive payments that this allows them to bid more competitively. If this continues to be the case, this may be an effective way for State agencies, which are legally bound by the competitive bidding system, to put more work in the hands of highly qualified contractors.

Base Course Air Voids and Surface Course Air Voids and Total Thickness

Compute Composite Quality Measure (PD*):

$$\text{PD}^* = 0.807\ \text{PD}_{VOIDS} + 0.669\ \text{PD}_{THICK} - 0.00476\ \text{PD}_{VOIDS} \times \text{PD}_{THICK}$$

(If base course, or no total thickness requirement,
use $\text{PD}_{THICK} = 10$ to compute PD*.)

Compute Percent Payment Adjustment (*PPA*):

<u>Mainline, Ramps, and New Shoulders</u>

$\text{PD}^* < 40$: $PPA = 10 - 0.67\ \text{PD}^*$
$\text{PD}^* \geq 40$: $PPA = 116 - 3.32\ \text{PD}^*$ (Minimum = –100 percent)

RETEST: $\text{PD}^* \geq 40$
REJECT: $\text{PD}^* \geq 65$

<u>Existing Shoulders</u>

Same as new shoulders except multiply computed *PPA* by 0.5.
(Minimum = –100 percent)

Surface Smoothness (Interim Procedure Until New Device Selected)

Compute Payment Adjustment (*PA*, expressed in units of $\$/m^2$):

$\text{PD}_{SMOOTH} < 2.0$: $PA = 0.34 - 0.26\ \text{PD}_{SMOOTH}$
$\text{PD}_{SMOOTH} \geq 2.0$: $PA = 0.72 - 0.45\ \text{PD}_{SMOOTH}$

RETEST: $\text{PD}_{SMOOTH} \geq 2.0$
REJECT: $\text{PD}_{SMOOTH} \geq 3.5$

Figure 44. Outline of NJDOT Superpave Acceptance Procedure

Case Study 3: Prototype Smoothness Specification

This case study describes the development of a new acceptance procedure for HMAC pavement smoothness based on International Roughness Index (IRI) and illustrates some of the preparatory steps that must be taken, and assumptions that must be made, when performance models are not readily available.

As described in case study 2, an interim acceptance procedure for pavement smoothness had been developed as part of the latest version of the NJDOT Superpave specification. The interim procedure is based on measurements obtained with the rolling straightedge, a type of device the NJDOT has used for many years. By this method, smoothness was judged based on PDL, defined as the percent of the length of the pavement lot having deviations exceeding 3.175 mm in 3.048 m, computed from dye marks made on the pavement as the device is pushed at walking speed.

The rolling straightedge has the advantage of being relatively inexpensive and its operation is easily understood. From this standpoint it was well suited to use for acceptance by highway agencies, and also by contractors for QC purposes. However, it also has several drawbacks in that data collection is very slow and labor intensive, lanes being measured must be closed to traffic, and operators are exposed to fast moving traffic in adjacent lanes. More recently, questions about its accuracy and precision have been raised, and literature is available indicating that measurements made with a 3.048 m straightedge are incapable of being a good measure of pavement smoothness because they are insensitive to some of the longer wavelengths that are important in terms of vehicle dynamics, particularly in regard to trucks and larger vehicles that are known to be the major contributors to pavement damage. (26) Consequently, it was decided to switch to a new measure of smoothness, the IRI, a measurement that is computed directly from the longitudinal profile of the pavement and that is designed to be sensitive to vehicle dynamics.

An interim acceptance procedure for smoothness based on IRI has been developed and, for the initial phase–in period, it has been recommended that any lots subject to rejection or substantial payment reduction be retested with the rolling straightedge to make the final determination of acceptance. This will provide a grace period in which both the NJDOT and the construction industry can become familiar with the IRI as the new measure of smoothness. It is recognized that IRI and PDL do not correlate well because they measure different things, but it is believed that an interim procedure including the rolling straightedge will be more acceptable to both industry and NJDOT personnel.

Risks Associated with New Procedure

Because the final decision will occasionally be based on the rolling straightedge during the phase-in period, there is the risk that pavements with poor IRI levels may be accepted, but that is exactly the situation now with acceptance based on PDL. There is also the likelihood that, on occasion, some pavements that are penalized based on IRI may appear to ride reasonably well in a passenger car and, if checked, may also have relatively low levels of PDL as measured with the rolling straightedge. This will most likely happen when the roughness consists of longer wavelengths that tend to affect larger and heavier vehicles, and it is a natural condition that must be expected to occur from time to time.

Measurement Device

Since the NJDOT currently owns an ARAN that has recently been equipped with laser sensors to measure pavement profile, this vehicle will be used as the measurement device, at least initially. Several tests have been run to confirm that the measurements are sufficiently repeatable, and it was also discovered that the degree of repeatability can be improved by using a method that automatically triggers the collection of data at a specified starting point. Eventually, it will probably be necessary to purchase additional equipment dedicated solely to this function.

Statistical Measures of Quality

Since PD has been used as the statistical quality measure for other construction specifications, and it is desired to control both the mean level and the variability of IRI, the new specification will also be based on PD. Based on recent literature, (26) plus an analysis of IRI data obtained from New Jersey highways, it was determined that an appropriate upper limit for the prototype specification would be an IRI value of 1.26 m/km (or 80 inches/mile), combined with a typical value for the AQL of PD = 10. Since larger IRI values represent rougher pavement, this is a one–sided specification with a single upper limit, and the PD above this limit will be designated as PD_{80}.

Unlike some statistical estimation procedures, for which the amount but not the location of the defective material is determined, the measurement of IRI produces a record of the pavement profile that makes it possible to identify the specific locations of rough areas. The availability of this information makes it possible to require corrective action, such as the selective grinding of particularly rough spots. Consequently, it was decided to define another quality measure, N_{100}, which represents a count of the individual sections of pavement having an IRI of 100 inches/mile, or more.

The acceptance procedure will be based on both quality measures—PD_{80} and N_{100}. For the prototype specification, N_{100} will be used only as one of the triggers to determine when the rolling straightedge must be brought out to make the final determination of acceptance. Eventually, it may be used to determine when corrective action (selective grinding) will be required. A flow chart of the proposed procedure is shown in figure 45.

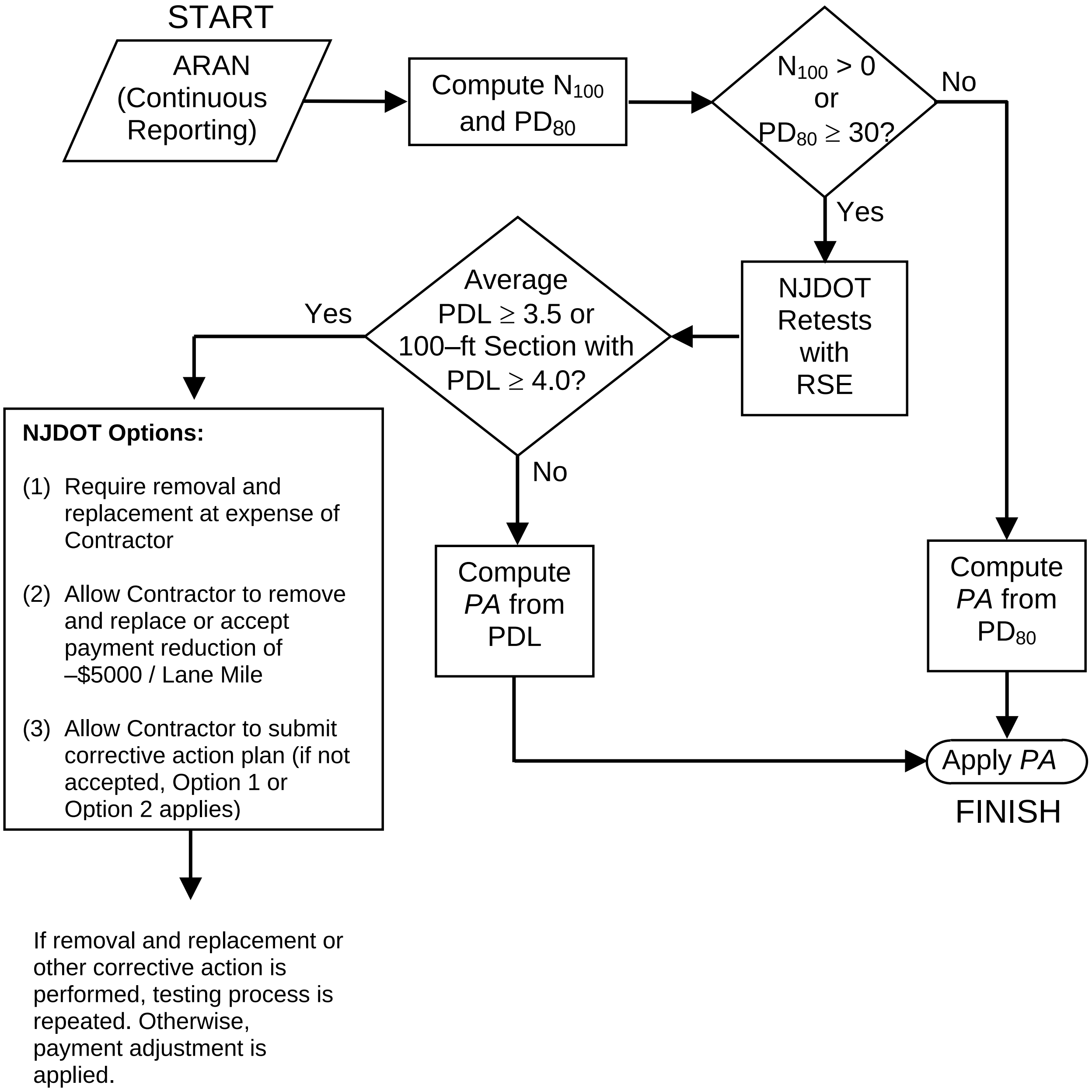


Figure 45. Proposed Smoothness Acceptance Procedure Based on IRI

Bonus Provision

The need for the use of a bonus provision to assure fairness with acceptance procedures based on PD (or PWL) has been discussed earlier. What is less clear is whether such a clause is appropriate with an “identify and correct” type of acceptance procedure. However, because an extremely high level of smoothness would be expected to extend pavement life, and a high level of initial quality would also tend to expedite the construction process, it was decided that some degree of bonus was justifiable. It has also been observed that many other agencies have chosen to include bonus provisions as part of their smoothness acceptance procedures. The presence of a bonus provision will have the further advantage of promoting better cooperation from the construction industry, and it may also allow better contractors to bid more competitively.

Lot Definition

In the past, the NJDOT has used specific starting and stopping stations to define smoothness lots, usually linked to time of construction (day’s production, etc.). However, with the use of a measuring device that can be driven down the road at close to traffic speeds, it is recognized that it will be more practical to define acceptance lots as the entire project in one direction, or possibly a single lane for the entire project in one direction, rather than attempting to identify precise starting and stopping points for each day’s production.

Payment Schedules

Similarly, it will be simpler (and is consistent with the LCC basis) if the payment adjustments are expressed in terms of dollars per lane–mile (or possibly dollars per square yard) rather than relating them to the bid price of the HMAC. (The interim acceptance procedure based on the rolling straightedge used with the latest Superpave pilot projects described in case study 2 expressed the payment adjustments in units of $\$/m^2$.)

However, to apply LCC to determine appropriate payment levels, it is necessary to have at least an approximate performance model to predict expected pavement life as a function of various levels of quality received. Since no such model based on IRI was readily available, it was necessary to rely on engineering knowledge and experience to develop a preliminary performance model.

As discussed in chapter 6, and developed further in appendix L, a particularly useful performance model is the exponential form given by equation 38,

$$EXPLIF = Ae^{-B(\mathrm{PD})^C} \qquad (38)$$

where $EXPLIF$ = expected life in years.
PD = the percent defective of the quality characteristic.
A, B, C = constants to be determined.

This model form is practical for several reasons. It tends to produce a sigmoidal (“S”) shape (although this may not always occur) that is well-suited for many quality characteristics. This property recognizes that there is often a point of diminishing returns, both for extremely good quality and extremely poor quality. For example, performance tends to improve relatively rapidly for increasing quality within the region in which most quality estimates tend to fall but, for extremely high levels of quality, the additional improvement often is only marginal. The same effect is usually observed for extremely poor quality, primarily because expected life is limited at zero and, as quality continues to decrease, expected life approaches the horizontal axis asymptotically. Consequently, this model provides a measure of realism that is not always present with other model forms.

Another advantage of the exponential model given by equation 38 is that it only requires three "known" points to determine the unknown coefficients—*A*, *B*, and *C*. In most cases, it will be possible to identify three points that are sufficiently well known that the model can be specified reasonably accurately. Points that are typically used are the origin (at which PD = 0 and coefficient *A* is the maximum expected life), the AQL (at which the expected life equals the design life), and the RQL (at which expected life is dramatically reduced, but usually is not zero).

Very similar reasoning was used for the development of the IRI performance model. The primary determining point is the AQL at $PD_{80} = 10$, at which it is assumed that the typical service life of 10 years for a resurfacing will be achieved. Next, based on a consensus of experienced engineers, it was decided that the best possible quality level, represented by $PD_{80} = 0$, might extend the life of a typical resurfacing to about 12 years. Finally, instead of attempting to estimate the expected life at the RQL (which, at this point, is yet to be defined), it was decided to use the poorest possible quality level of $PD_{80} = 100$. Based on experience with very rough pavements, plus recent literature on IRI measurements, it was believed that such a pavement would not be immediately repaired in many cases and, consequently, would not have an expected life of zero.(26) However, this clearly is an extremely poor level of quality, and it was decided to assume an expected life of two years for purposes of developing the preliminary model. This produced the performance matrix given by table 31.

Table 31. Performance Matrix for Smoothness Specification Based on IRI

PD_{80}	***EXPLIF*, in years**
0	12
10	10
100	2

The next step is to use the data from this performance matrix to determine the unknown model coefficients. By taking logarithms of both sides of equation 38, and substituting the values for PD_{80} and *EXPLIF* from table 31, it is possible to write three equations in three unknowns, and solve to obtain the performance model given by equation 39. (As a check, it is easy to demonstrate that, when the PD_{80} values from table 31 are entered, this equation returns the appropriate values for *EXPLIF*.)

$$EXPLIF = 12e^{-0.0186(\mathrm{PD}_{80})^{0.992}} \tag{39}$$

Once the performance model has been obtained, a method must be found to use the estimates of expected life it provides and convert them into appropriate levels of adjusted payment. To do this, the LCC equation derived in appendix I, which is repeated below as equation 40, can be used.

$$PAYADJ = \frac{C(R^D - R^E)}{(1 - R^O)} \tag{40}$$

where: *PAYADJ* = appropriate payment adjustment for new pavement or overlay (same units as *C*).

C = present total cost of resurfacing (typical value = \$23.92/m^2 (\$20/square yard)).

D = design life of pavement or initial overlay (typically 20 years for new pavement, 10 years for overlay).

E = expected life of pavement or overlay (variable).

O = expected life of successive overlays (typically 10 years).

R = (1 + *INF*) / (1 + *INT*).

INF = long–term annual inflation rate in decimal form (typically 0.04).

INT = long–term annual interest rate in decimal form (typically 0.08).

The next step is to use equations 39 and 40 to develop table 32, relating PD_{80} to *EXPLIF* and *PAYADJ*. These results are then plotted in figure 46.

Table 32. PD Related to *EXPLIF* and *PAYADJ*

PD_{80}	**EXPLIF, in years** (computed with equation 39)	**PAYADJ, \$ / Lane Mile** (computed with equation 40)
0	12.0	+22,300
10	10.0	0
30	7.0	–36,800
50	4.9	–65,200
70	3.4	–86,900
90	2.4	–102,000

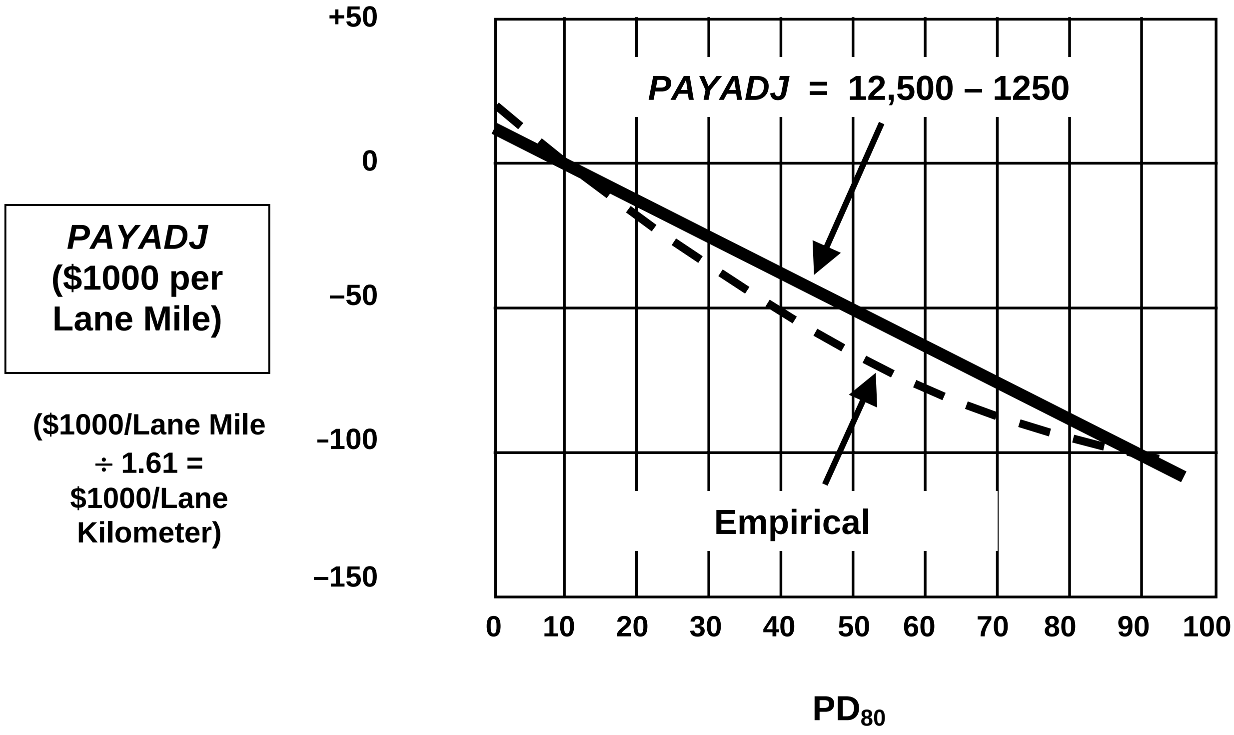


Figure 46. Proposed Smoothness Acceptance Procedure Based on IRI

The empirical relationship obtained from equations 39 and 40 is plotted as a dashed line in figure 46. Since this relationship is nearly linear, it is approximated by the linear payment schedule that is given by equation 41, and plotted in figure 46. This represents the level of adjusted payment, in $ per lane–mile, that is justified by LCC analysis applied to the tentative performance model given by equation 39.

$$PAYADJ \ = \ 12{,}500 \ - \ 1250 \ \mathrm{PD}_{80} \qquad (41)$$

An agency could choose to use any payment schedule that is less severe than this, as long as it provided sufficient incentive to produce the desired level of quality. For example, if management felt that the maximum bonus of +$12,500 per lane mile awarded by equation 41 was excessive, and wished to cap it at +$5000 per lane mile instead, then a somewhat shallower payment equation would be necessary. Since this shallower payment equation must go through the point $\mathrm{PD}_{80} = 10$, $PAYADJ = 0$, in addition to the point $\mathrm{PD}_{80} = 0$, $PAYADJ = +\$5000$, its intercept is +$5000 and its slope is (5000 – 0) / (0 – 10) = –500, leading to equation 42 as just one of many alternate payment equations that might be used.

$$PAYADJ \ = \ 5000 \ - \ 500 \ \mathrm{PD}_{80} \qquad (42)$$

For reasonably good levels of IRI quality, payment schedules such as equations 41 or 42 would apply. However, if either $N_{100} > 0$ or $PD_{80} \geq 30$ based on the initial tests, then a retest with the rolling straightedge is required, and the payment schedule given by equation 43, where PDL is percent defective length, applies (subject to the options listed on the flow chart in figure 45). Equation 43 is an interim payment schedule chosen to be very close to an existing schedule used with the rolling straightedge, and will no longer apply when the rolling straightedge is eventually phased out.)

$$PAYADJ \ = \ 3000 \ - \ 2308 \ \mathrm{PDL} \tag{43}$$

It can be seen by inspection that equation 43 pays a maximum bonus of *PAYADJ* = +$3000/lane mile at the best possible quality level of PDL = 0, and produces a payment adjustment of zero (100 percent payment) at the AQL of PDL = 1.3 for this measurement device. At the RQL value of PDL = 3.5, the payment reduction will be approximately –$5000/lane mile, provided the RQL provision in figure 45 is not enforced.

Implementation and Future Modifications

At the time of this writing, the final details of the specification are being worked out in preparation for testing on a series of pilot projects during the next construction season. As with other trial specifications, it is expected that the payment schedules will be reduced, perhaps by as much as 50 percent, for the initial pilot projects. Provided these projects are successful, it is planned to eventually exclude the use of the rolling straightedge before implementing the new specification widely on all projects.

Various modifications of this acceptance procedure are anticipated, both near–term and long–term. Depending upon the level of quality received, and on the ability of the construction industry to meet these requirements, some minor changes of the acceptance procedure may be appropriate in the near–term. Normally, the second phase would be to restore the full amount of the payment schedules, and then schedule additional pilot projects.

Long–term, the tendency will be to continue to “raise the bar” by specifying higher levels of quality as the construction industry gradually adapts to the new requirements and shows that they can consistently meet them. Also, since the performance model upon which the acceptance procedure is based is only tentative, it will be necessary to track performance so that, when some of the pavements constructed under this specification eventually require resurfacing, it will be possible to assess the adequacy of the model.

APPENDIX A

Bibliography of Selected QA–Related Publications

"23 CFR Part 637," *Quality Assurance Procedures for Construction*, Federal Highway Administration, Federal Register, July 1995.

AASHTO Implementation Manual for Quality Assurance, American Association of State Highway and Transportation Officials, AASHTO Quality Construction Task Force, February 1996.

AASHTO Quality Assurance Guide Specification, American Association of State Highway and Transportation Officials, AASHTO Quality Construction Task Force, February 1996.

"Acceptance Sampling Plans for Highway Construction," *American Association of State Highway and Transportation Officials Materials, Part I Specifications: R–9*, American Association of State Highway and Transportation Officials, 1984.

Afferton, K.C., Freidenrich, J. and Weed, R.M., "Managing Quality: Time for a National Policy," *Transportation Research Record 1340*, Transportation Research Board, National Research Council, Washington, DC, 1992, pp. 3–39.

Amirkhanian, S.N., Burati, J.L., and Mirchandani, H.C., "Effect of Testing Variability on Contractor Payment for Asphalt Pavements," *Journal of Construction Engineering and Management*, vol. 120, No. 3, September 1994, pp. 579–592.

Anderson, D.A., et al., "Framework for Development of Performance–Related Specifications for Hot–Mix Asphaltic Concrete," *National Cooperative Highway Research Program Report 332*, Transportation Research Board, Washington, DC, January 1990.

Benson, P.E., "Comparison of End Result and Method Specifications for Managing Quality," *Transportation Research Record 1491*, Transportation Research Board, National Research Council, Washington, DC, 1995, pp. 3–10.

Benson, P.E., "Performance Review of a Quality Control/Quality Assurance Specification for Asphalt Concrete," *Transportation Research Record 1654*, Transportation Research Board, National Research Council, Washington, DC, 1999, pp. 88–94.

Burati, J.L. and Willenbrock, J.H., "Acceptance Criteria for Bituminous Surface Course on Civil Airport Pavements," *FAA-RD-79-89*, Federal Aviation Administration, October 1979.

Burati, J.L. and Hughes, C.S., "Highway Materials Engineering Module I: Materials Control and Acceptance—Quality Assurance," *National Highway Institute Course No. 13123*, Federal Highway Administration, Washington, DC, 2001.

Chamberlin, W.P., “Performance–Related Specifications,” *NCHRP Synthesis 212*, Transportation Research Board, Washington, DC, 1995.

Darter, M.I., et al., “Development of Prototype Performance–Related Specification for Concrete Pavements,” *Transportation Research Record 1544*, Transportation Research Board, National Research Council, Washington, DC, 1996, pp. 81–90.

Darter, M.I., et al., “Performance–Related Specifications for Concrete Pavements, Volume I,” *Report No. FHWA–RD–93–042*, Federal Highway Administration, Washington, DC, November 1993.

Darter, M.I., et al., “Performance–Related Specifications for Concrete Pavements, Volume II,” *Report No. FHWA–RD–93–043*, Federal Highway Administration, Washington, DC, November 1993.

Deming, W.E., *Out of the Crisis*, MIT CAES, Cambridge, 1986.

Dobrowolski, J., and Bressette, T., “Development of Quality Control/Quality Assurance Specifications by Using Statistical Quality Assurance for Asphalt Concrete Pavements in California,” *Transportation Research Record 1632*, Transportation Research Board, National Research Council, Washington, DC, 1998, pp. 13–21.

Douglas, K.D., et al., “Evaluation of Quality Control/Quality Assurance Implementation for Asphalt Concrete Specifications in California,” *Transportation Research Record 1654*, Transportation Research Board, National Research Council, Washington, DC, 1999, pp. 95–101.

Elliott, R.P., and Yanjun, Q., “Analysis of Contractor Pay Adjustment Schedule Using Simulation,” *Transportation Research Record 1544*, Transportation Research Board, National Research Council, Washington, DC, 1996, pp. 109–115.

Gentry, C., and Yrjanson, W.A., “Specifications for Quality Control: A Case Study,” *Transportation Research Record 1126*, Transportation Research Board, National Research Council, Washington, DC, 1987, pp. 37–41.

“Glossary of Highway Quality Assurance Terms,” *Transportation Research Circular Number E-C037*, Transportation Research Board, Washington, DC, April 2002.

Grant, E.L. and Leavenworth, R.S., *Statistical Quality Control, Seventh Ed.*, ISBN 0–07–043555–3, McGraw–Hill, January 1996.

Hoerner, T.E., and Darter, M.I., “Guide to Developing Performance–Related Specifications for PCC Pavements, Volume I: Practical Guide, Final Report, and Appendix A,” *Report No. FHWA–RD–98–155*, Federal Highway Administration, Washington, DC, February 1999.

Hoerner, T.E., "Guide to Developing Performance–Related Specifications for PCC Pavements, Volume IV: Appendix for PaveSpec 2.0 User Guide," *Report No. FHWA–RD–99–059*, Federal Highway Administration, Washington, DC, February 1999.

Hoerner, T.E., et al., "Improved Prediction Models for PCC Pavement Performance–Related Specifications, Volume I: Final Report," *Report No. FHWA–RD–00–130*, Federal Highway Administration, Washington, DC, August 2000.

Hoerner, T.E., and Darter, M.I, "Improved Prediction Models for PCC Pavement Performance–Related Specifications, Volume II: PaveSpec 3.0 User Guide," *Report No. FHWA–RD–00–131*, Federal Highway Administration, Washington, DC, August 2000.

Hughes, C.S., "Incentive and Disincentive Specification for Asphalt Concrete Density," *Transportation Research Record 986*, Transportation Research Board, National Research Council, Washington, DC, 1984, pp. 38–42.

Hughes, C.S., "Variability in Highway Pavement Construction," *NCHRP Synthesis 232*, National Cooperative Highway Research Program, Transportation Research Board, National Research Council, Washington, DC, 1996**.**

Hughes, C.S., Simpson, A.L., Cominsky, R., Maser, K., and Wilson, T., "Measurement and Specification of Construction Quality, Volume I," *Final Report*, Federal Highway Administration, FHWA–RD–98–077, May 1998.

Irick, P.E., "A Conceptual Framework for the Development of a Performance–Related Materials and Construction Specification," *Transportation Research Record 1126*, Transportation Research Board, National Research Council, Washington, DC, 1987, pp. 1–27.

Irick, P.E., et al., "Development of Performance–Related Specifications for Portland Cement Pavement Construction," *Report No. FHWA–RD–89–211*, Federal Highway Administration, Washington, DC, May 1990.

Kandahl, P.S., et al., "Development and Implementation of Statistically Based End–Result Specifications for Hot–Mix Asphalt in Pennsylvania," *Transportation Research Record 1389*, Transportation Research Board, National Research Council, Washington, DC, 1993, pp. 9–16.

Manual on Presentation of Data and Control Chart Analysis, Sixth Ed., American Society for Testing and Materials Committee E–11 on Quality and Statistics, Philadelphia, PA, 1992.

The Memory Jogger II ^{TM}, ISBN 1–879364–44–1, GOAL/QPC, Salem, NH, January 1994.

Moore, R.M., et al., "Overview of Pay–Adjustment Factors for Asphalt Concrete Mixtures," *Transportation Research Record 821*, Transportation Research Board, National Research Council, Washington, DC, 1981, pp. 49–56.

“Quality Assurance,” *NCHRP Synthesis 65*, Transportation Research Board, Washington, DC, 1979.

Schmitt, R.L., et al., “Summary of Current Quality Control/Quality Assurance Practices for Hot–Mix Asphalt Construction,” *Transportation Research Record 1632,* Transportation Research Board, National Research Council, Washington, DC, 1998, pp. 22–31.

Seeds, S. B., et al., “Development of Performance–Related Specifications for Hot–Mix Asphalt Pavements Through WesTrack,” *Transportation Research Record 1575*, Transportation Research Board, National Research Council, Washington, DC, 1997, pp. 85–91.

“Statistically Oriented End–Result Specifications,” *NCHRP Synthesis 38*, Transportation Research Board, Washington, DC, 1976.

Vlatas, D.A. and Smith, R.E., “Implications of Life–Cycle Performance Specifications,” *Transportation Research Record 1215*, Transportation Research Board, National Research Council, Washington, DC, 1989, pp. 25–30.

Weed, R.M., “Stratified Random Sampling From a Discrete Population,” *Transportation Research Record 792,* Transportation Research Board, National Research Council, Washington, DC, 1981, pp. 41–44.

Weed, R.M., “Method to Establish Pay Schedules for Rigid Pavement,” *Transportation Research Record 885*, Transportation Research Board, National Research Council, Washington, DC, 1982, pp. 18–24.

Weed, R.M., “Development of Multicharacteristic Acceptance Procedures for Rigid Pavement,” *Transportation Research Record 885*, Transportation Research Board, National Research Council, Washington, DC, 1982, pp. 25–36.

Weed, R.M., “Adjusted Pay Schedules: New Concepts and Provisions,” *Transportation Research Record 986*, Transportation Research Board, National Research Council, Washington, DC, 1984, pp. 32–38.

Weed, R.M., “Computer–Assisted Random Sampling,” *Transportation Research Record 1034,* Transportation Research Board, National Research Council, Washington, DC, 1985, pp. 140–152.

Weed, R.M., “Revision of a Flawed Acceptance Standard,” *Transportation Research Record 1056,* Transportation Research Board, National Research Council, Washington, DC, 1986, pp. 21–35.

Weed, R.M., “Composite Pay Equations: General Approach,” *Transportation Research Record 1465,* Transportation Research Board, National Research Council, Washington, DC, 1994, pp. 9–15.

Weed, R.M., “Development of Air Voids Specification for Bituminous Concrete,” *Transportation Research Record 1491*, Transportation Research Board, National Research Council, Washington, DC, 1995, pp. 33–39.

Weed, R.M., “OCPLOT: Program to Generate Operating Characteristic Curves for Statistical Construction Specifications,” *Transportation Research Record 1491*, Transportation Research Board, National Research Council, Washington, DC, 1995, pp. 18–26.

Weed, R.M., “Quality Assurance Software for the Personal Computer,” *Transportation Research Record 1544*, Transportation Research Board, National Research Council, Washington, DC, 1996, pp. 116–124.

Weed, R.M., “A Rational Method for Relating As–Built Quality to Pavement Performance and Value,” *Transportation Research Record 1632,* Transportation Research Board, National Research Council, Washington, DC, 1998, pp. 31–39.

Weed, R.M., “Practical Framework for Performance–Related Specifications,” *Transportation Research Record 1654*, Transportation Research Board, National Research Council, Washington, DC, 1999, pp. 81–87.

Weed, R.M., “Development of Composite Quality Measures,” *Transportation Research Record 1712,* Transportation Research Board, National Research Council, Washington, DC, 2000, pp. 103–108.

Weed, R.M., “Method to Model Performance Relationships and Pay Schedules,” *Transportation Research Record 1712,* Transportation Research Board, National Research Council, Washington, DC, 2000, pp. 117–124.

Weed, R.M., “New Jersey’s Superpave Specification: The Next Generation,” *Transportation Research Record 1761,* Transportation Research Board, National Research Council, Washington, DC, 2001, pp. 10–14.

Weed, R.M., “Derivation of Equation for Cost of Premature Pavement Failure,” *Transportation Research Record 1761,* Transportation Research Board, National Research Council, Washington, DC, 2001, pp. 93–96.

Welborn, J.Y., “State–of–the–Art in Asphalt Pavement Specifications,” *Report No. FHWA/RD–84/075*, Federal Highway Administration, Washington, DC, July 1984.

Wheeler, D.J., *Understanding Variation: The Key to Managing Chaos*, ISBN 0–945320–35–3, Statistical Process Control Press, June 1993.

APPENDIX B

Example QC Plan for HMAC (16)

This HMAC QC Plan is submitted for Project xxxxx, item 401.

HMAC Plant Production

1. Plant (Enter make, type and location of plant).
   1.1 Frequency of plant inspection, calibration, verification of calibration, and any plant certification. (Include documentation).

2. Personnel.
   2.1 Enter name, certification number and telephone number of employee responsible for QC.
   2.2 Enter name, and certification no. of sampling & testing technician.
   2.3 Enter name and telephone number of employee responsible for making plant production changes when necessary as a result of QC data.

3. Enter the type and mix design identification of mixes to be used in the contract.

4. Prior to production the JMF for each type of mix included for use on the contract will be submitted to the agency. Only materials from sources acceptable to the agency will be used in the mix design.

5. During mix production operations, QC tests will be performed at or exceeding the minimum frequency in the attached schedule.

6. All testing and evaluation will be completed within _____ hours of sampling and all documentation will be completed and submitted to the agency on approved processing forms within ______ hours or production will be halted until these item are current.

7. Material found to be noncomplying shall not be incorporated into the work.

8. Appropriate agency personnel will be notified at least 24 hours before the scheduled work is to begin.

9. In the event QC test data indicates that noncomplying material has been incorporated into the work, the agency will be notified immediately and a plan will be submitted for agency approval, identifying, the defective material and its disposition.

HMAC Field Operations QC

1. Personnel.
  1.1 Enter name, qualification, and telephone number of employee directing the field operations.
  1.2 Enter name, and qualification, of employee responsible for insuring that all items of work will comply with agency specifications.
  1.3 Enter name and certification number (if applicable) of employee responsible for sampling and testing.

2. During placement operations of the asphalt concrete pavement, QC tests will be performed at or exceeding the minimum frequency in the attached schedule.

3. All testing and evaluation will be completed within the time limits specified by the contractor or the work will be halted.

4. Material found to be noncomplying will not be incorporated into the roadway.

5. Notification will be given to appropriate agency personnel at least 24 hours before work is scheduled to begin.

6. In the event that test data or inspections indicate that noncomplying material has been incorporated into the work, the agency will be notified immediately and a plan will be submitted for agency approval, identifying the detective material and its disposition.

7. Indicate method to be followed to prevent segregation and visual pavement deformities.

Example Table for Frequency and Documentation of Sampling and Testing Asphalt Concrete

Hot–Mix Asphalt Concrete		
Test	**Frequency**	**Documentation**
Combined cold feed gradation	As required to control production	Forms as required by contract
Moisture content	As required to control production	Forms as required by contract
Asphalt content	As required to control operation	Forms as required by contract
Tack/prime, if used	As required to control production	Daily
Correlation of nuclear asphalt gauge	1 per mix or project	Forms as required by contract
Face fracture on crushed gravel, if used	Lot size to be determined. 5 per lot	Forms as required by contract
Temperature of mix	As required to control production	Daily
Temperature of base or air	As needed	Daily
Temperature of mat	As required to control operation	Daily
Density	Lot size to be designated. 5 per lot	Forms as required by contract
Skid resistance	As required to control operation	Forms as required by contract
Smoothness	As required to control operation	Forms as required by contract
Thickness	Lot size to be designated. 5 per lot	Forms as required by contract
Pavement application rate	As needed	Daily
Test data distributed	Within 24 hours of sampling	Forms as required by contract

APPENDIX C

Example QC Plan for Structural PCC [(16)]

1. The following outlines the contractor's plan for insuring QC in accordance with contractual requirements for the production, shipment, placement, and curing of PCC which will be used in the listed contract items.
   1.1 Item xxxx—Bridge.
   1.2 Item xxxx—Headwalls.
   1.3 Item xxxx—Curb & gutter.

2. Personnel.
   2.1 Qualified sampling, testing and inspection personnel who are certified through agency accepted programs will be used.
   2.2 Technicians (enter name and certification number of technician(s)).
   2.3 The QC Liaison with the agency (enter name, certification number and telephone number).
   2.4 The QC Supervisor (enter name, qualification and telephone number).

3. Recognized statistical concepts for material data analysis as allowed by agency will be used.

4. Documentation of QC activities will be available at (enter the location at which the documentation will be maintained).

5. A copy of typical report forms used documenting QC work is attached. (The agency should attach appropriate forms.) These forms will include:
   5.1 Identification—Name of material, equipment, or process being evaluated.
   5.2 Location of sample procurement or inspection or testing.
   5.3 Type of test or inspection.
   5.4 Name of inspector, sampler, tester and reviewer.
   5.5 Results—Observation of inspection or test value result.
   5.6 Analysis of Acceptability.
   5.7 Action taken and results.
   5.8 Signature of QC supervisor.

6. Typical control charts used for documenting QC work are attached. (The agency should attach appropriate forms.) These charts will include the following data:
   6.1 Identification—Name of material, equipment, or process being evaluated.
   6.2 Results—Date, and report number of data used on chart.
   6.3 Upper & lower control limits.
   6.4 Analysis of acceptability.
   6.5 Action taken and results.
   6.6 Signature of responsible QC lead employee.

7. Other documentation used for QC analysis will include the following:
   7.1 Batch tickets.
   7.2 Other records not addressed above.

8. In the event QC test data indicates that noncomplying material has been incorporated into the work, the agency will be notified immediately and a plan will be submitted for agency concurrence, to identify the defective material and determine its disposition.

9. Mix Design(s), (enter class and identification of mix designs to be used on items in the contract).

Example Table for Frequency of Sampling and Testing for QC of PCC Components

Fine and Coarse Aggregates		
Item	**Test**	**Frequency**
Gradation	AASHTO T 27	As required for control

APPENDIX D

Example QC Plan for Rigid Pavement (16)

1. The following outlines (enter contractor's name) plan for insuring QC in accordance with contractual requirements for the production, shipment, placement, and curing of Item xxx, Rigid Pavement, which will be used on Project xxxx.

2. Personnel.
  2.1 Qualified sampling, testing and inspection personnel who are certified through agency accepted programs will be used.
  2.2 Technicians (enter name and certification number of technician(s)).
  2.3 The QC Liaison with the agency (enter name, certification number and telephone number).
  2.4 The QC Supervisor (enter name, qualification and telephone number).

3. Recognized statistical concepts for material data analysis as allowed by agency specifications will be used. These are (list references or detail the concepts to be followed).

4. Documentation of QC activities will be available at (enter the location at which the documentation will be maintained).

5. A copy of typical report forms used documenting QC work are attached. [The agency should attach appropriate forms.] These forms will include:
  5.1 Identification—Name of material, equipment, or process being evaluated.
  5.2 Location of sampling, inspection or testing.
  5.3 Type of test or inspection.
  5.4 Name of inspector, sampler, tester and reviewer.
  5.5 Results—Observation of inspection or test value result.
  5.6 Analysis of acceptability.
  5.7 Action taken and results.
  5.8 Signature of QC supervisor.

6. Typical control charts used for documenting QC work are attached. [The agency should attach appropriate forms.] These charts will include the following, data:
  6.1 Identification—Name of material, equipment, or process being, evaluated.
  6.2 Results—Date, and report number of data used on chart.
  6.3 Upper and lower control limits.
  6.4 Analysis of acceptability.
  6.5 Action taken and results.
  6.6 Signature of responsible QC lead employee.

7. Other documentation used for QC analysis will include the following:

  7.1 Batch tickets.

  7.2 Other records not addressed above.

8. If QC test data indicates that noncomplying material has been inadvertently been incorporated into the work, the agency will be notified immediately and a plan will be submitted for review and concurrence, identifying the defective material and determine its disposition.

9. Mix Design(s), (enter class and identification of mix designs to be used on items in the contract). Supporting, test data attached.

Example Tables for Frequency of Sampling and Testing for QC of Rigid Pavement Components

Fine and Coarse Aggregates		
Item	**Test**	**Frequency**
Gradation	AASHTO T 27	As required for control

Example Tables for Frequency of QC Inspection, Sampling, and Testing for Production of PCC Pavement

Plant and Trucks		
Item Identification	**Item to Check**	**Frequency**
Mixer Blades	Wear and alignment	Daily
Scales	Static check	Weekly
Belts & rollers	Wear and alignment	Weekly
Calibration Gauges	Verify calibration	Monthly

Concrete Production—Stockpile Maintenance		
Item Identification	**Test**	**Frequency**
Moisture	Determine moisture content	As needed to control production
Contamination	Visual	Daily
Segregation	Visual	Daily

Example Tables for Frequency of QC Inspection, Sampling, and Testing for Production of PCC Pavement

Concrete Production—Concrete		
Item	**Test**	**Frequency**
Yield	AASHTO T 121	As required to control production
Sump	AASHTO T 119	As required to control production
Air	AASHTO T 152 or T 196	As required to control production
Strength	AASHTO T 22 or T 276	As required to control production
Temperature	ASTM C 1077	As required to control production
Mixing time	Visual	Weekly/daily per batch size

Example List of Requirements for QC of Transport and Placement of PCC Paving Mix

10. Tests on concrete.
  10.1 Slump—As required to control the product.
  10.2 Air content—As required to control the product.
  10.3 Compressive strength—as required to control the product.
  10.4 Placement and consolidation.
    10.4.1 Formwork—Tightness, bracing, etc., visual.
    10.4.2 Vibration—Visual per specification.
    10.4.3 Rate of pour—Visual per specification.
    10.4.4 Replacement equipment—Condition visual.
    10.4:5 Finishing, equipment—Condition visual.
    10.4.6 Staffing—number for production, ability visual.
    10.4.7 Texturing—Visual per specification.
    10.4.8 Thickness—Measure.
    10.4.9 Rebar cover—Measure.
    10.4.10 Smoothness—Measure.
    10.4.11 Concrete surface conditions and geometry—Visual inspection.
    10.4.12 Curing—Visual inspection.
    10.4.13 Weather Protection—Visual inspection.

APPENDIX E

Flowchart of 23 CFR 637B

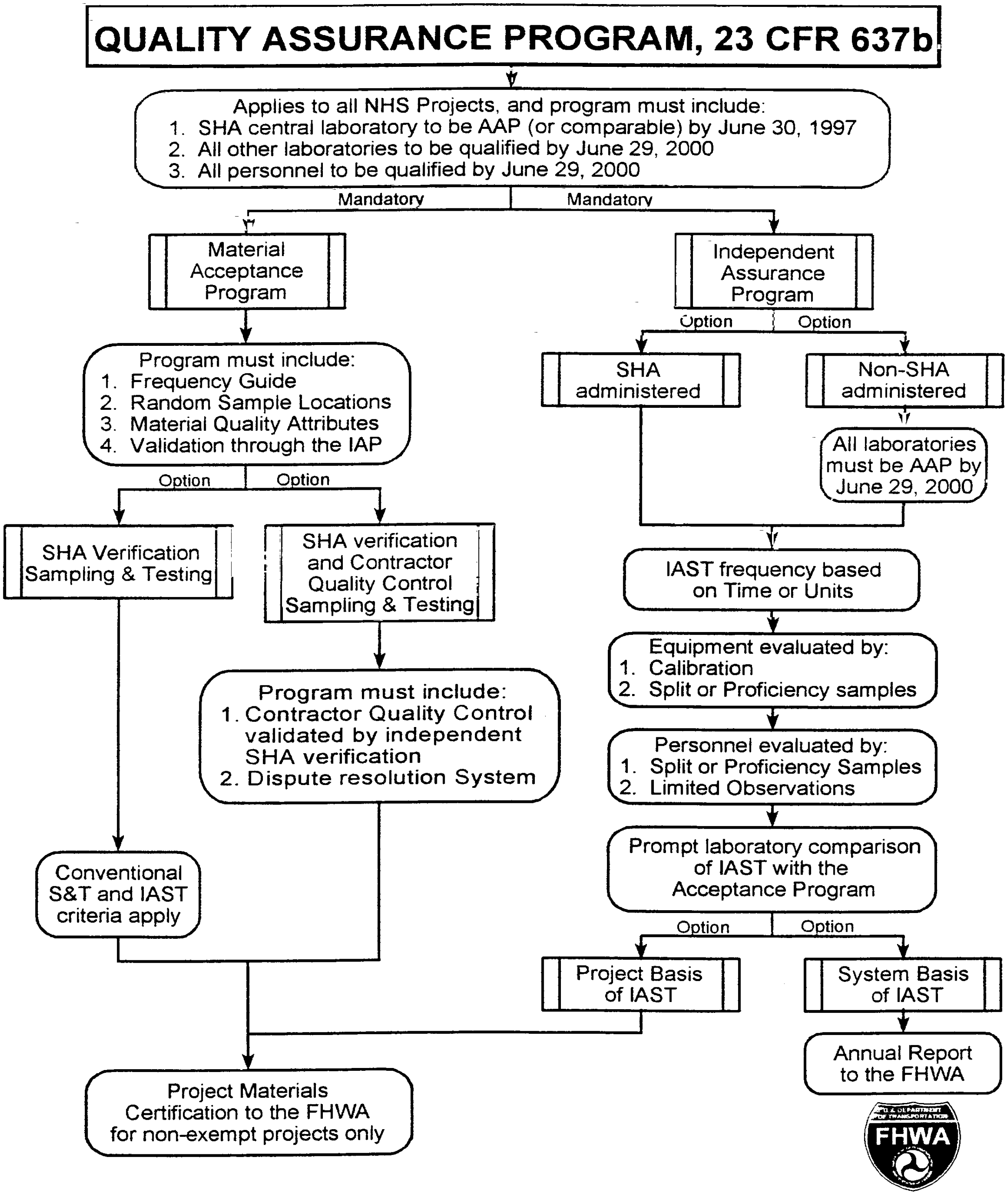


Figure 47. Flowchart of 23 CFR 637B

APPENDIX F

F–test and *t*–test Method for Comparing Two Sets of Data

Introduction

In comparing two sets of data, such as contractor and agency test results, what is involved is two hypothesis tests, where the H_0 for each test is that the data sets are from the same population. In other words, the null hypotheses are that the variabilities of the two data sets are equal, for the *F*–test, and that the means of the two data sets are equal, for the *t*–test.

When comparing two data sets, it is important to compare both the means and the variances. A different test is used for each of these comparisons. The ***F*–test** provides a method for comparing the **variances** (standard deviation squared) of the two sets of data. Differences in **means** are assessed by the ***t*–test**. Construction processes and material properties usually follow a normal distribution. For normal distributions, the ratios of variances follow an *F*–distribution, while the means of relatively small samples follow a *t*–distribution. Hypothesis tests for equal variances and means can therefore be conducted using these distributions.

For samples from the same normal population, the statistic *F*, which is the ratio of the two sample variances, has a sampling distribution called the *F*–distribution. Tables are available for the *F*–distribution just like they are for the normal distribution. For process verification testing, the *F*–test is based on the ratio of the sample variance of the contractor's test results, s_c^2, and the sample variance of the agency's test results, s_a^2.

Similarly, the *t*–statistic and the *t*–test can be used to test whether the sample mean of the contractor's test results, $\overline{X}_c$**,** and that of the agency's test results $\overline{X}_a$, came from populations with the same mean.

The equations for the *F*–test and *t*–test are presented conceptually in the following sections, but it is recommended that a computer program be used in practice to perform the calculations. Spreadsheet programs, such as Microsoft® Excel, have both *F*–tests and *t*–tests. Agencies may also wish to develop their own computer packages. Also, the program DATATEST, which was developed for FHWA Demonstration Project 89, is demonstrated at the end of this appendix. (18)

When comparing contractor and agency samples, it is important that **random sampling** was used when obtaining the samples. Also, because sources of variability influence the population parameters, the two sets of test results must have been sampled over the **same time period**, and the **same sampling and testing procedures** must have been used. If it is determined that a significant difference is likely between either the variances or the means, the source of the difference should be identified. The identification of a difference is just that, i.e., notice that a difference exits. The reason for the difference must still be determined.

Before comparing contractor and agency samples, **a level of significance, α, must be selected**. While α values of 0.10, 0.05, and 0.01 are common, many agencies select a value of 0.01 to minimize the likelihood of incorrectly concluding that the results are different when they actually came from the same population. However, it should be recognized that selecting a low α value reduces the chance of detecting a real difference when one actually exists.

F–test for Sample Variances

Since the values used for the *t*–test are dependent upon whether or not the variances are assumed equal for the two data sets, it is necessary to **test the variances before the means**. The intent is to determine whether the difference in the variability of the contractor's tests and the agency's tests is larger than might be expected by chance if they came from the same population. It does not matter which variance is larger. After comparing the *F*–test results, one of the following will be concluded:

- The two sets of data have different variances because the difference between the two sets of test results is greater than is likely to occur from chance if their variances are actually equal.
- There is no reason to believe the variances are different because the difference is not so great as to be unlikely to have occurred from chance if the variances are actually equal.

Steps Involved in the *F*–test

The first step is to compute the variance for the contractor's tests, s_c^2, and the agency's tests, s_a^2. Then use the simple ratio equation to compute *F*, where $F = s_c^2 / s_a^2$ or $F = s_a^2 / s_c^2$. *Always use the larger of the variances in the numerator so the ratio will be greater than 1.*

Next, choose α, the level of significance for the test. For this discussion $\alpha = 0.01$ is used.

The next step is to determine the critical *F* value, F_{crit}, from the *F*–table (see table 35 at the end of this appendix) for the α level of significance chosen, and using the degrees of freedom $(n - 1)$ associated with each set of test results. Thus, the degrees of freedom associated with the contractor's variance, s_c^2, is $(n_c - 1)$ and the degrees of freedom associated with the agency's variance, s_a^2, is $(n_a - 1)$. The values in this *F*–table are tabulated to test if there is a difference (either larger or smaller) between the two variance estimates. This is known as a two–sided or two–tailed test. Care must be taken when using other tables of the *F*–distribution, since they are usually based on a one–tailed test, i.e., testing whether one variance is larger than another is. This means that the F_{crit} values in table 35 are the same values that would be listed at the 99.5 percentile (even though the 99.0 percentile would normally be associated with $\alpha = 0.01$) for a one–sided test.

Once the value for F_{crit} is determined from the table (making sure the appropriate degrees of freedom for the numerator and denominator are used), if $F \geq F_{crit}$, then decide that the two sets of tests have significantly different variabilities. If $F < F_{crit}$ then decide that there is no reason to believe that the variabilities are significantly different.

F–test Example Problem 1

A contractor has run 12 asphalt content tests and the agency has run 6 tests over the same period of time using the same sampling and testing procedure. The results are shown below. Based on their variabilities, is it likely that the tests came from the same population?

Table 33. Asphalt Content Tests

Contractor Tests	Agency Tests
6.41	5.42
6.23	5.78
6.08	6.23
6.55	5.38
6.11	5.62
5.97	5.79
6.28	—
6.07	—
5.92	—
5.76	—
6.06	—
5.71	—
$\overline{X} = 6.10$	$\overline{X} = 5.70$
$s_c^2 = 0.061$	$s_c^2 = 0.097$

Use the *F*–test to determine whether or not to assume the variance of the contractor's tests differs from the variance of the agency's tests.

Step 1. Compute the variance, s^2, for each set of tests.

$$s_c^2 = 0.061 \qquad s_a^2 = 0.097 \qquad (44, 45)$$

Step 2. Compute *F*:

$$F = \frac{s_a^2}{s_c^2} = \frac{0.097}{0.061} = 1.59 \qquad (46)$$

Step 3. Determine F_{crit} from the *F*–distribution table making sure to use the correct degrees of freedom for the numerator ($n_a - 1 = 6 - 1 = 5$) and the denominator ($n_c - 1 = 12 - 1 = 11$). From table 35, $F_{crit} = 6.42$.

Conclusion: Since $F < F_{crit}$ (i.e., $1.59 < 6.42$), there is no reason to believe that the two sets of data have different variabilities. That is, they could have come from the same population.

F–test Example Problem 2

A contractor has run 10 air void tests from cores and the agency has run 5 air void tests over the same period of time using the same sampling and testing procedure. The results are shown below. Based on their variabilities, is it likely that the tests came from the same population?

Table 34. Air Void Tests

Contractor Tests	Agency Tests
6.42	7.52
7.18	11.38
5.04	9.20
4.56	5.32
7.12	3.18
7.98	—
6.32	—
6.08	—
5.92	—
5.78	—
$\overline{X} = 6.24$	$\overline{X} = 7.32$
$s_c^2 = 1.036$	$s_c^2 = 10.299$

Step 1. Compute the variance, s^2, for each set of tests.

$s_c^2 = 1.036$ $s_a^2 = 10.299$ (47, 48)

Step 2. Compute *F*:

$$F = \frac{s_a^2}{s_c^2} = \frac{10.299}{1.036} = 9.94 \quad (49)$$

Step 3. Determine F_{crit} from the *F*–distribution table making sure to use the correct degrees of freedom for the numerator ($n_a - 1 = 5 - 1 = 4$) and the denominator ($n_c - 1 = 10 - 1 = 9$). From table 35, $F_{crit} = 7.96$.

Conclusion: Since $F > F_{crit}$ (i.e., 9.94 > 7.96), it is unlikely that the two data sets came from the same population. Therefore, conclude that the contractor and agency results are different.

t–test for Sample Means

Once the variances have been tested and assumed to be either equal or not equal, the means of the test results can be tested to determine whether they differ from one another or can be assumed to be equal. The desire is to determine whether it is reasonable to assume that the contractor's tests came from the same population as the agency's tests. A ***t*–test** is used to **compare the sample means**. Two approaches for the *t*–test are necessary. If the sample variances are assumed equal (*F*–test example problem 1 above), then the *t*–test is conducted based on the two samples using a **pooled** estimate for the variance and the **pooled** degrees of freedom. This approach is *t*–test example 1 described below. If the sample variances are assumed to be different (*F*–test example problem 2 above), then the *t*–test is conducted using the individual sample variances, the individual sample sizes, and the **effective** degrees of freedom (estimated from the sample variances and sample sizes). This approach is *t*–test example 2 below.

In either of the two cases discussed in the previous paragraph, one of the following decisions is made:

- The two sets of data have different means because the difference in the sample means is greater than is likely to occur from chance if their means are actually equal.
- There is no reason to believe that the means are different because the difference in the sample means is not so great as to be unlikely to have occurred from chance if the means are actually equal.

Conceptually, for the *t*–test in which the **sample variances are equal**, the equation used to calculate the *t*–value divides the difference between two means by the pooled standard deviation. The pooled standard deviation is the square root of the pooled variance that is the weighted average of the two variances, using the degrees of freedom for each sample as the weighting factor. (Again, conceptually, this is similar to the *Z*–equation in which the difference between the mean and a point of interest is expressed in standard deviation units. But because small sample sizes are used, the *t*–distribution is used.)

To determine the critical *t* value, t_{crit}, against which the computed *t*–value is compared, it is necessary to select the level of significance, α. Again, a value of $\alpha = 0.01$ is recommended. Next, the critical *t*–value, t_{crit}, is obtained from the *t*–table (see table 36 at the end of this appendix) for the pooled degrees of freedom. The pooled degrees of freedom for the case where the sample variances are assumed equal are $(n_c + n_a - 2)$. If $t \geq t_{crit}$, then decide that the two sets of tests have significantly different means. If $t < t_{crit}$, then decide that there is no reason to believe the means are significantly different.

t–test Example Problem 1: Sample Variances Assumed to Be Equal.

Use *F*–test example problem 1 above in which a contractor has run 12 asphalt content tests and the agency has run 6 tests over the same period of time using the same sampling and testing procedures. Based on their means, is it likely that the tests came from the same population?

Use the *t*–test for the case of equal variances (determined above in *F*–test example problem 1) to determine whether or not to assume the mean of the contractor's tests differs from the mean of the agency's tests.

In *F*–test example problem 1, it was determined that $s_c^2 = 0.061$ and $s_a^2 = 0.097$.

Step 1. Compute the sample mean, $\overline{X}$, for each set of tests.

$$\overline{X}_c = 6.10 \qquad \overline{X}_a = 5.70 \tag{50, 51}$$

Step 2. Compute the pooled variance, s_p^2, using the sample variances from above.

$$s_p^2 = \frac{s_c^2(n_c - 1) + s_a^2(n_a - 1)}{n_c + n_a - 2} \tag{52}$$

$$s_p^2 = \frac{0.061(12-1) + 0.097(6-1)}{12 + 6 - 2} = 0.072$$

Step 3. Compute the *t*–statistic, *t*, using the equation for equal variances.

$$t = \frac{\left|\overline{X}_c - \overline{X}_a\right|}{\sqrt{\frac{s_p^2}{n_c} + \frac{s_p^2}{n_a}}} \tag{53}$$

$$t = \frac{|6.10 - 5.70|}{\sqrt{\frac{0.072}{12} + \frac{0.072}{6}}} = 2.981$$

Step 4. Determine the critical *t* value, t_{crit}, for the pooled degrees of freedom.

Degrees of freedom = ($n_c + n_a - 2$) = (12 + 6 – 2) = 16.

From table 36, for $\alpha = 0.01$ and 16 degrees of freedom, $t_{crit} = 2.921$.

Conclusion: Since 2.981 > 2.921, we reject the null hypothesis, and assume that the sample means are not equal. We therefore assume that they came from different populations. We therefore conclude that it is unlikely (but not impossible) that the contractor and agency test results represent the same process. In other words, the agency tests do not verify the contractor tests.

t–test Example Problem 2: Sample Variances Assumed to be Different

The *F*–test example problem 2 above in which a contractor has run 10 air void tests from cores and the agency has run 5 tests over the same period of time using the same sampling and testing procedure is used. Based on their means, is it likely that the tests came from the same population?

In *F*–test example problem 2, it was determined that $s_c^2 = 1.036$ and $s_a^2 = 10.299$.

Step 1. Compute the mean, $\overline{X}$, for each set of tests.

$$\overline{X}_c = 6.24 \qquad\qquad \overline{X}_a = 7.32 \tag{54, 55}$$

Step 2. Compute the *t*–statistic, *t*, using the equation for unequal variances.

$$t = \frac{\left|\overline{X}_c - \overline{X}_a\right|}{\sqrt{\dfrac{s_c^2}{n_c} + \dfrac{s_a^2}{n_a}}} \tag{56}$$

$$t = \frac{\left|6.24 - 7.32\right|}{\sqrt{\dfrac{1.036}{10} + \dfrac{10.299}{5}}} = 0.734$$

Step 3. Determine the critical *t* value, t_{crit}, for the effective degrees of freedom, f'.

$$f' = \frac{\left(\dfrac{s_c^2}{n_c} + \dfrac{s_a^2}{n_a}\right)^2}{\left[\dfrac{\left(\dfrac{s_c^2}{n_c}\right)^2}{n_c + 1} + \dfrac{\left(\dfrac{s_a^2}{n_a}\right)^2}{n_a + 1}\right]} - 2 \tag{57}$$

$$f' = \frac{\left(\dfrac{1.036}{10} + \dfrac{10.299}{5}\right)^2}{\left[\dfrac{\left(\dfrac{1.036}{10}\right)^2}{10 + 1} + \dfrac{\left(\dfrac{10.299}{5}\right)^2}{5 + 1}\right]} - 2 = 4.61 \rightarrow 5$$

The calculated value for effective degrees of freedom is rounded to the closest integer in this example. The critical value could also be obtained by interpolation or by truncating to the lowest integer. This equation is an approximation and there is not a universally accepted method for arriving at the effective degrees of freedom. In general, rounding to a smaller value for degrees of freedom gives a larger critical value, thereby making it less likely to reject the null hypothesis of equal means.

Note that the value for effective degrees of freedom is less than would have been used if the variances had been assumed to be equal.

From the *t*–table, table 36, for $\alpha = 0.01$ and 5 degrees of freedom, $t_{crit} = 4.032$.

Conclusion: Since $0.734 < 4.032$, there is no reason to reject the assumption that the means are equal. Therefore, we assume that it is possible (but not certain) that they came from the same population.

Note: The difference in sample means is much greater in this example (7.32 – 6.24 = 1.08) than in the previous example (6.10 – 5.70 = 0.40). However, in the previous example it was concluded that the means were different, while in this example it was not concluded that the means were different. The larger ratio of variance values in this example is the reason that it was not possible to conclude that the means were different.

Computer Programs for the *F*–test and *t*–test Calculations

As can be seen from the example problems, the required computations can be quite complex and time consuming. This introduces the possibility of human error.

Using Microsoft Excel.

As noted above, spreadsheet programs such as Microsoft Excel often have built–in functions for conducting both *F*–tests and *t*–tests. These tests can be performed by anyone with a basic knowledge regarding how to use spreadsheet functions. Excel has a function for conducting *F*–tests. Excel can also conduct paired *t*–tests, as well as two–sample *t*–tests for the cases of both equal and unequal variances.

To illustrate the use of spreadsheets for conducting *F*–tests and *t*–tests, Excel was used to compare the data sets used in Example Problem 1 above. The following paragraphs show the steps necessary in using Excel for these calculations.

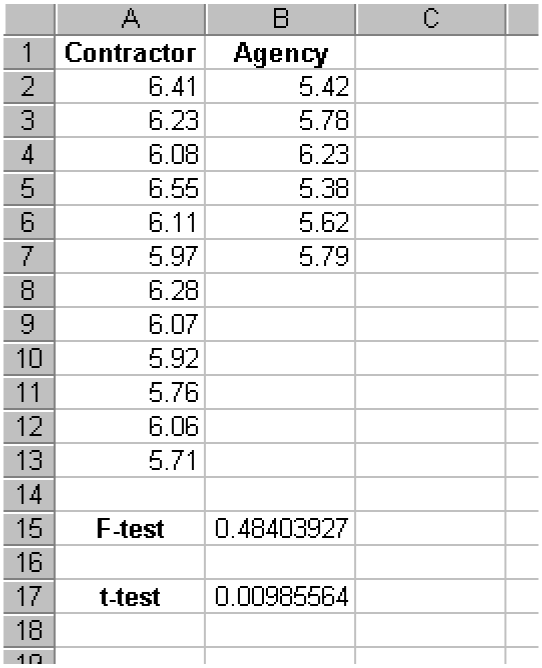
The first step is to input the contractor and agency data into two different columns in Excel. The data for this example are shown in figure 48.

The *F*–test is then conducted before the *t*–test. This is done by using the Excel function

FTEST(array1,array2)

where: array1 is the array representing one set of data
array2 is the array representing the other set of data.

For the example in figure 48, the contractor data are in array1, and it is input as A2:A13, while the agency data are in array2 and it is input as B2:B7. The function that is entered into cell B15 is therefore =FTEST(A2:A13,B2:B7).



	A	B	C
1	**Contractor**	**Agency**	
2	6.41	5.42	
3	6.23	5.78	
4	6.08	6.23	
5	6.55	5.38	
6	6.11	5.62	
7	5.97	5.79	
8	6.28		
9	6.07		
10	5.92		
11	5.76		
12	6.06		
13	5.71		
14			
15	**F-test**	0.48403927	
16			
17	**t-test**	0.00985564	
18			

Figure 48. Excel Results for Data from Example Problem 1

The test that is conducted by Excel is a <u>one–sided</u> *F*–test. The value that is displayed in cell B15 is the probability of getting an *F*–value as large as the one for these data sets if the two data sets have the same variance. In other words, the lower the probability value returned by this function, the less likely it is that the two sets of data have the same variance. For example, if the level of significance for the test were selected as 0.05, for a one–tailed test you would reject the assumption of equal variances whenever the probability value that is returned by the function is less than 0.05.

To compare the results of function FTEST with the critical values in table 35, which is based on a <u>two–sided</u> *F*–test and $\alpha = 0.01$ therefore, you would reject the assumption of equal variances whenever the Excel FTEST function returned a probability value less than 0.005. Figure 48 shows that for the example data a probability value of 0.484 is returned by the FTEST function. Therefore, the conclusion would be to assume that the variances are equal.

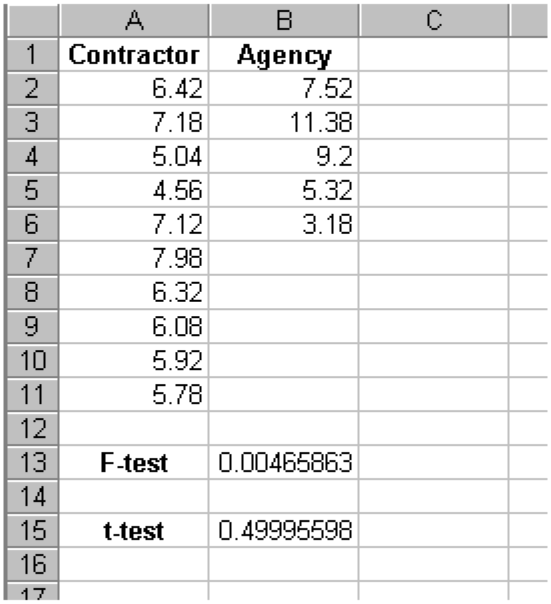
Once the results of the *F*–test are known, the *t*–test can then be conducted using the Excel function

TTEST(array1,array2,tails,type)

Where: array1 is the array representing one set of data.
array2 is the array representing the other set of data.
tails is either 1 for a one–sided test or 2 for a two–sided test.
type is 1 for a paired t–test, 2 for an equal variance t–test, and 3 for an unequal variance t–test.

For the example in figure 48, the contractor data are in array1, and it is input as A2:A13, while the agency data are in array2 and it is input as B2:B7. Since a two–tailed is desired, tails is input as 2, and, since from the *F*–test the variances were assumed to be equal, type is input as 2. The function that is entered into cell B17 is therefore =TTEST(A2:A13,B2:B7,2,2). Figure 48 shows that for the example data a probability value of 0.00986 is returned by the TTEST function. Therefore, at the α = 0.01 level of significance, the conclusion would be to assume that the means are not equal since the probability value is less than 0.01.

Similarly, Excel can be used to perform the *F*–test and *t*–test on the data sets from Example Problem 2 above. This is illustrated in figure 49.



	A	B	C
1	**Contractor**	**Agency**	
2	6.42	7.52	
3	7.18	11.38	
4	5.04	9.2	
5	4.56	5.32	
6	7.12	3.18	
7	7.98		
8	6.32		
9	6.08		
10	5.92		
11	5.78		
12			
13	**F-test**	0.00465863	
14			
15	**t-test**	0.49995598	
16			

Figure 49. Excel Results for Data from Example Problem 2

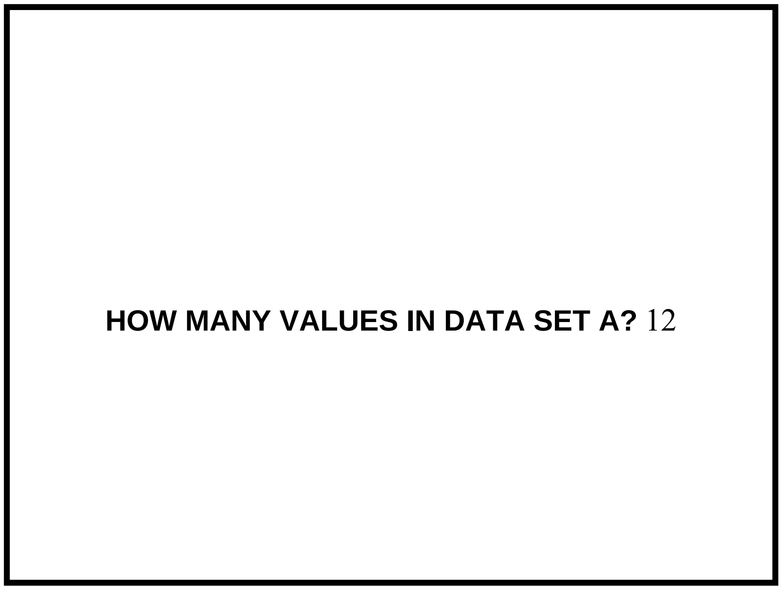
The results in figure 49 (see cell B13) indicate that the variances are assumed to be not equal. This means that the type input for the TTEST function will be 3, for an unequal variance *t*–test. The tails input will still be 2 for a two–tailed test. The results in figure 49 (see cell B15) indicate that the means are assumed to be equal since the probability in cell B15 is much greater than the level of significance of α = 0.01.

Using Program DATATEST

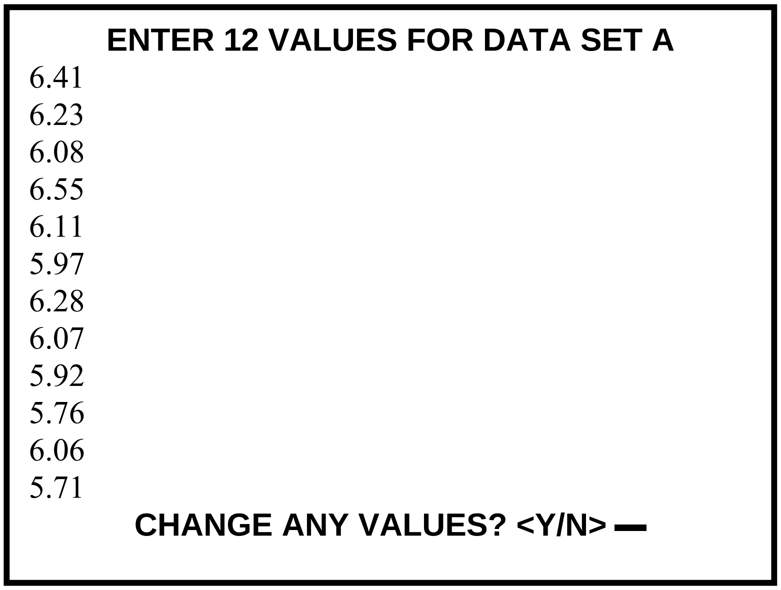
Another software program that can be used for performing *F*–test and *t*–test comparisons is the FHWA Demonstration Project No. 89 program DATATEST. (18) This program demonstrates how simply the *F*–tests and *t*–tests can be performed with a personal computer. To illustrate this, the DATATEST program was used to compare the data sets used in the example problems above. To illustrate the use of the program, the input and output screens for these examples are presented in the figures beginning on the next page.

DATATEST Screens for the Data from Example Problem 1

The program first asks for the number of values and then allows the user to input the values for the first set of data.

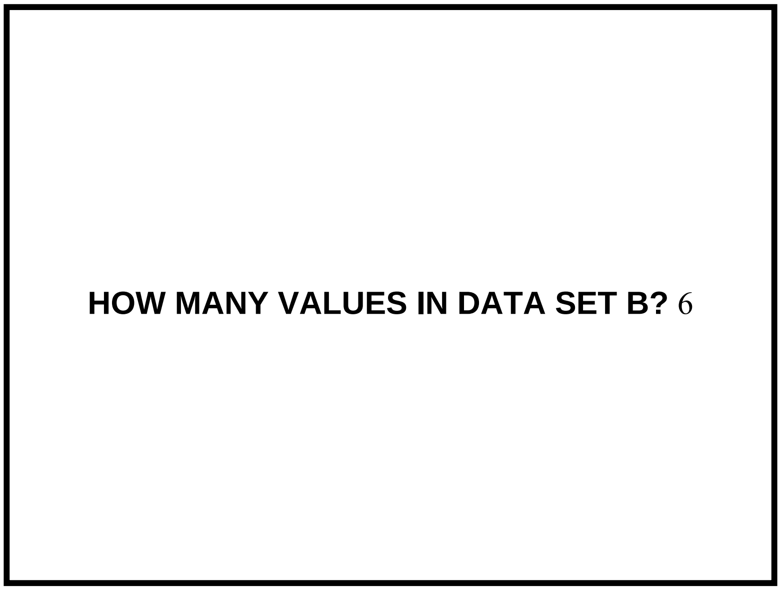


```
HOW MANY VALUES IN DATA SET A? 12
```

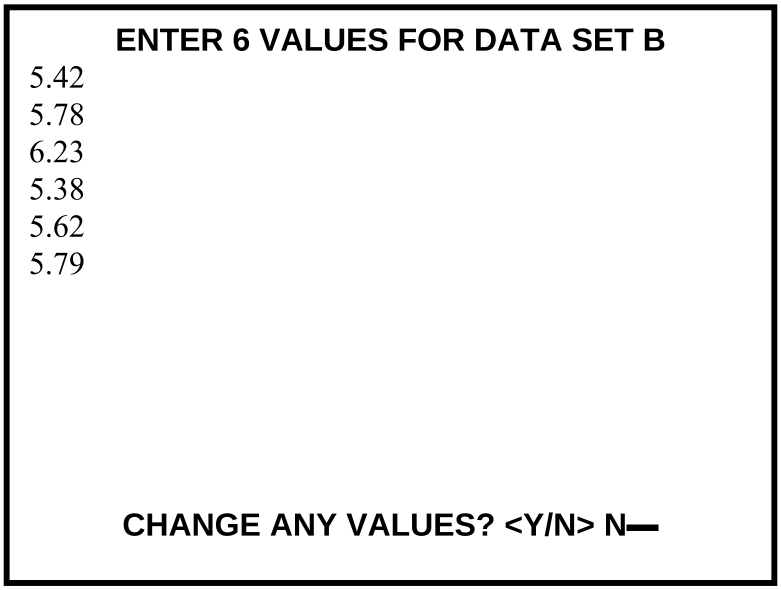


```
ENTER 12 VALUES FOR DATA SET A
6.41
6.23
6.08
6.55
6.11
5.97
6.28
6.07
5.92
5.76
6.06
5.71
CHANGE ANY VALUES? <Y/N> ▬
```

The program then asks for the number of values and then allows the user to input the values for the second set of data.



```
HOW MANY VALUES IN DATA SET B? 6
```



```
ENTER 6 VALUES FOR DATA SET B
5.42
5.78
6.23
5.38
5.62
5.79

CHANGE ANY VALUES? <Y/N> N▬
```

The program then asks the user to select a level of significance, α.

```
SELECT SIGNIFICANCE LEVEL
(ALPHA) TO BE USED FOR
F AND T TESTS

            (1) 0.01
            (2) 0.05
            (3) 0.10
          SELECTION?
```

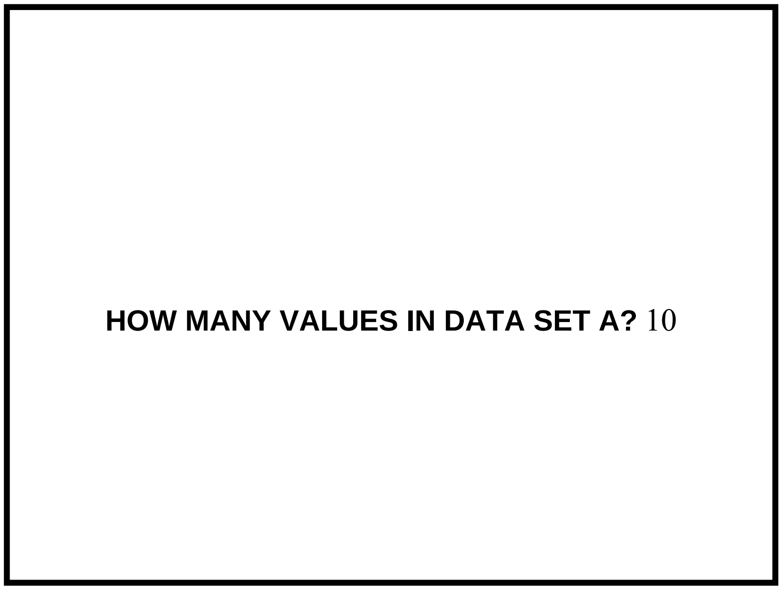
Finally, the program conducts the *F*–test and then, based on the *F*–test results, the appropriate form of the *t*–test, and displays the results.

```
DATA SET A                DATA SET B
   N = 12                    N = 6
   X̄ = 6.0958                X̄ = 5.7033
   S = .24637                S = .31066

COMPARE                   COMPARE
STANDARD                  SAMPLE
DEVIATIONS                MEANS
F(CALC) = 1.59            T(CALC) = 2.93
F(CRIT) = 6.42            T(CRIT) = 2.92

NO SIGNIFICANT            SIGNIFICANT
DIFFERENCE AT             DIFFERENCE AT
ALPHA = 0.01              ALPHA = 0.01

           Press any key to continue
```

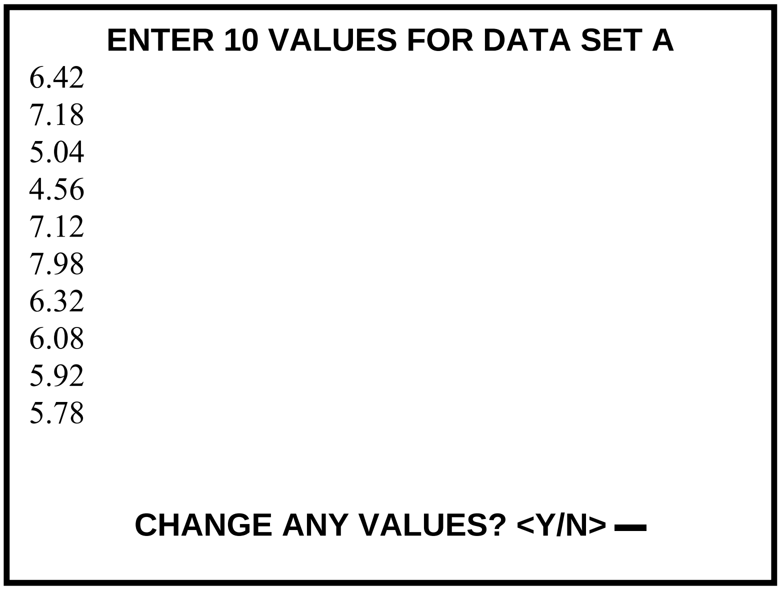
The values obtained by the DATATEST program are consistent with those calculated in Example Problem 1 above. The slight difference in the calculated *t*– value stems from the number of decimal places that are used in the computer's calculation. The results from the DATATEST program, i.e., the variances not assumed different and the means assumed different, are consistent with those from Example Problem 1.

DATATEST Screens for the Data from Example Problem 2

The program first asks for the number of values and then allows the user to input the values for the first set of data.

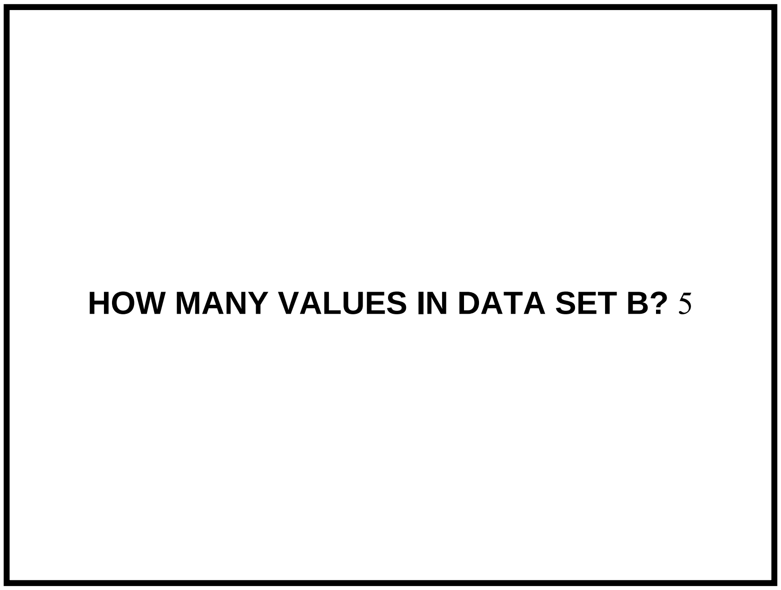


```
HOW MANY VALUES IN DATA SET A? 10
```

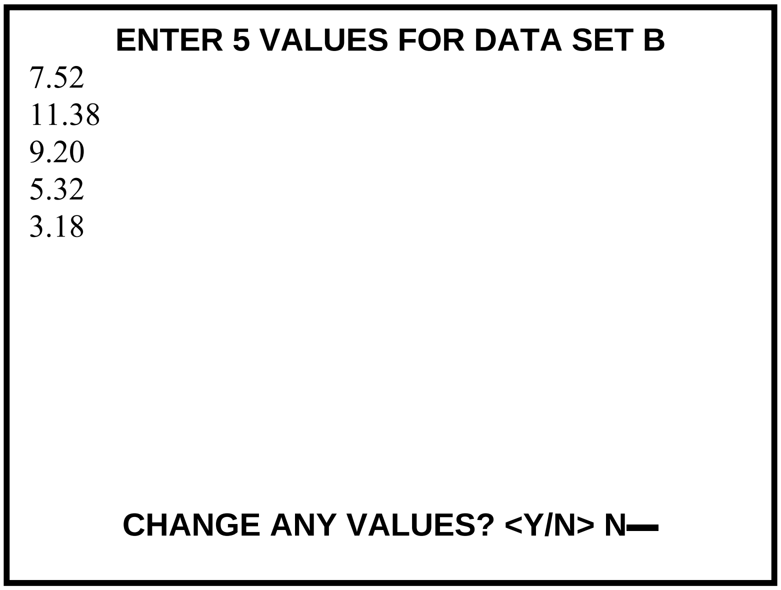


```
ENTER 10 VALUES FOR DATA SET A
6.42
7.18
5.04
4.56
7.12
7.98
6.32
6.08
5.92
5.78

CHANGE ANY VALUES? <Y/N> ▬
```

The program then asks for the number of values and then allows the user to input the values for the second set of data.

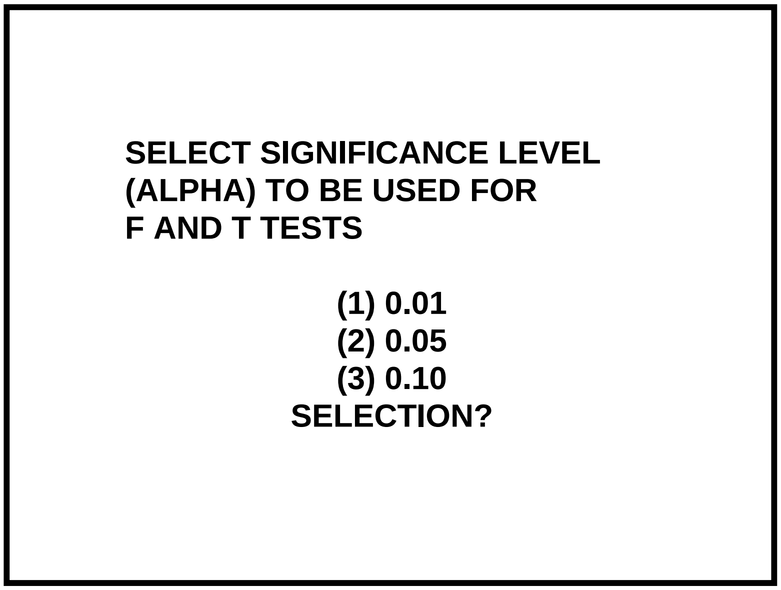


```
HOW MANY VALUES IN DATA SET B? 5
```



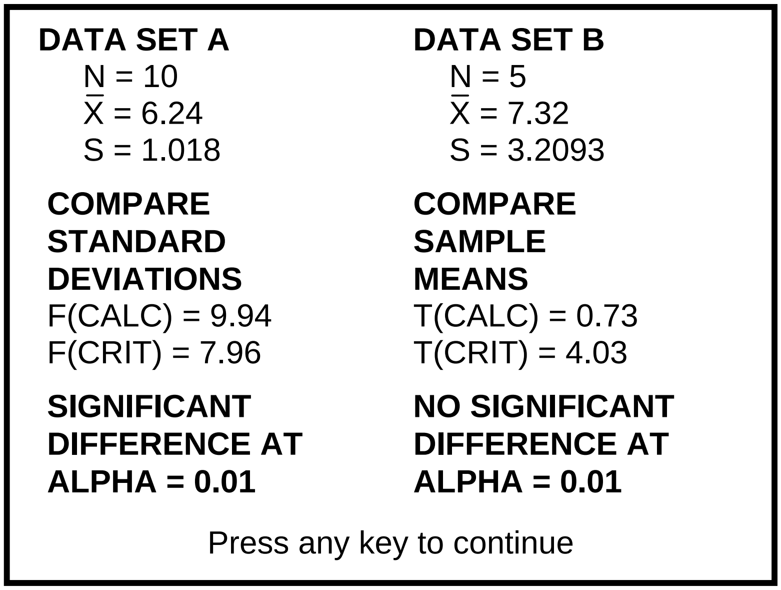
```
ENTER 5 VALUES FOR DATA SET B
7.52
11.38
9.20
5.32
3.18



CHANGE ANY VALUES? <Y/N> N▬
```

The program then asks the user to select a level of significance, α.



```
SELECT SIGNIFICANCE LEVEL
(ALPHA) TO BE USED FOR
F AND T TESTS

           (1) 0.01
           (2) 0.05
           (3) 0.10
          SELECTION?
```

Finally, the program conducts the *F*–test and then, based on the *F*–test results, the appropriate form of the *t*–test, and displays the results.



```
DATA SET A                DATA SET B
   N = 10                    N = 5
   X̄ = 6.24                  X̄ = 7.32
   S = 1.018                 S = 3.2093

 COMPARE                  COMPARE
 STANDARD                 SAMPLE
 DEVIATIONS               MEANS
 F(CALC) = 9.94           T(CALC) = 0.73
 F(CRIT) = 7.96           T(CRIT) = 4.03

 SIGNIFICANT              NO SIGNIFICANT
 DIFFERENCE AT            DIFFERENCE AT
 ALPHA = 0.01             ALPHA = 0.01

          Press any key to continue
```

The values obtained by the DATATEST program are consistent with those calculated in Example Problem 2 above. The results from the DATATEST program, i.e., the variances assumed different and the means not assumed different, are consistent with those from Example Problem 2.

Table 35. Critical Values, F_{crit}, for the F–test for a Level of Significance, $\alpha = 0.01$[1]

DEGREES OF FREEDOM FOR NUMERATOR

DEGREES OF FREEDOM FOR DENOMINATOR

	1	2	3	4	5	6	7	8	9	10	11	12
1	16200	20000	21600	22500	23100	23400	23700	23900	24100	24200	24300	24400
2	198	199	199	199	199	199	199	199	199	199	199	199
3	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.5	43.4
4	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.8	20.7
5	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.5	13.4
6	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	10.1	10.0
7	16.2	12.4	10.9	10.0	9.52	9.16	8.89	8.68	8.51	8.38	8.27	8.18
8	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.10	7.01
9	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.31	6.23
10	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.75	5.66
11	12.2	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.32	5.24
12	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.99	4.91
15	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.33	4.25
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.76	3.68
24	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.50	3.42
30	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.25	3.18
40	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	3.03	2.95
60	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.82	2.74
120	8.18	5.54	4.50	3.92	3.55	3.28	3.09	2.93	2.81	2.71	2.62	2.54
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.43	2.36

[1] **NOTE:** This is for a **two-tailed test** with the null and alternate hypotheses shown below:

$$H_0:\ s_c^2 = s_a^2$$
$$H_a:\ s_c^2 \neq s_a^2$$

Table 35. Critical Values, F_{crit}, for the F–test for a Level of Significance, $\alpha = 0.01$[1] *(continued)*

DEGREES OF FREEDOM FOR NUMERATOR

DEGREES OF FREEDOM FOR DENOMINATOR

	15	**20**	**24**	**30**	**40**	**50**	**60**	**100**	**120**	**200**	**500**	**∞**
1	24600	24800	24900	25000	25100	25200	25300	25300	25400	25400	25400	25500
2	199	199	199	199	199	199	199	199	199	199	199	200
3	43.1	42.8	42.6	42.5	42.3	42.2	42.1	42.0	42.0	41.9	41.9	41.8
4	20.4	20.2	20.0	19.9	19.8	19.7	19.6	19.5	19.5	19.4	19.4	19.3
5	13.1	12.9	12.8	12.7	12.5	12.5	12.4	12.3	12.3	12.2	12.2	12.1
6	9.81	9.59	9.47	9.36	9.24	9.17	9.12	9.03	9.00	8.95	8.91	8.88
7	7.97	7.75	7.65	7.53	7.42	7.35	7.31	7.22	7.19	7.15	7.10	7.08
8	6.81	6.61	6.50	6.40	6.29	6.22	6.18	6.09	6.06	6.02	5.98	5.95
9	6.03	5.83	5.73	5.62	5.52	5.45	5.41	5.32	5.30	5.26	5.21	5.19
10	5.47	5.27	5.17	5.07	4.97	4.90	4.86	4.77	4.75	4.71	4.67	4.64
11	5.05	4.86	4.76	4.65	4.55	4.49	4.45	4.36	4.34	4.29	4.25	4.23
12	4.72	4.53	4.43	4.33	4.23	4.17	4.12	4.04	4.01	3.97	3.93	3.90
15	4.07	3.88	3.79	3.69	3.59	3.52	3.48	3.39	3.37	3.33	3.29	3.26
20	3.50	3.32	3.22	3.12	3.02	2.96	2.92	2.83	2.81	2.76	2.72	2.69
24	3.25	3.06	2.97	2.87	2.77	2.70	2.66	2.57	2.55	2.50	2.46	2.43
30	3.01	2.82	2.73	2.63	2.52	2.46	2.42	2.32	2.30	2.25	2.21	2.18
40	2.78	2.60	2.50	2.40	2.30	2.23	2.18	2.09	2.06	2.01	1.96	1.93
60	2.57	2.39	2.29	2.19	2.08	2.01	1.96	1.86	1.83	1.78	1.73	1.69
120	2.37	2.19	2.09	1.98	1.87	1.80	1.75	1.64	1.61	1.54	1.48	1.43
∞	2.19	2.00	1.90	1.79	1.67	1.59	1.53	1.40	1.36	1.28	1.17	1.00

[1] **NOTE:** This is for a **two-tailed test** with the null and alternate hypotheses shown below:

$$H_0:\ s_c^2 = s_a^2$$
$$H_a:\ s_c^2 \neq s_a^2$$

Table 36. Critical Values, t_{crit}, for the t–test[1]

Degrees of Freedom	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
1	63.657	12.706	6.314
2	9.925	4.303	2.920
3	5.841	3.182	2.353
4	4.604	2.776	2.132
5	4.032	2.571	2.015
6	3.707	2.447	1.943
7	3.499	2.365	1.895
8	3.355	2.306	1.860
9	3.250	2.262	1.833
10	3.169	2.228	1.812
11	3.106	2.201	1.796
12	3.055	2.179	1.782
13	3.012	2.160	1.771
14	2.977	2.145	1.761
15	2.947	2.131	1.753
16	2.921	2.120	1.746
17	2.898	2.110	1.740
18	2.878	2.101	1.734
19	2.861	2.093	1.729
20	2.845	2.086	1.725
21	2.831	2.080	1.721
22	2.819	2.074	1.717
23	2.807	2.069	1.714
24	2.797	2.064	1.711
25	2.787	2.060	1.708
26	2.779	2.056	1.706
27	2.771	2.052	1.703
28	2.763	2.048	1.701
29	2.756	2.045	1.699
30	2.750	2.042	1.697
40	2.704	2.021	1.684
60	2.660	2.000	1.671
120	2.617	1.980	1.658
∞	2.576	1.960	1.645

---

[1] **NOTE:** This is for a two-tailed test with the null and alternate hypotheses shown below:

$$H_0: \overline{X}_c = \overline{X}_a$$
$$H_a: \overline{X}_c \neq \overline{X}_a$$

APPENDIX G

OC Curves for Various Methods for Comparing Two Sets of Data

As discussed in chapter 5, verification testing can be of two types: test method verification testing that is done on split samples, or process verification testing that is done on independent samples. The procedures are different for each of these types of verification testing.

OC Curves for Test Method Verification

In chapter 5, two methods were considered for test method verification of split samples: the D2S method, which compares the contractor and agency results from a single split sample, and the paired *t*–test, which compares contractor and agency results from a number of split samples. OC curves, which plot the probability of detecting a difference versus the actual difference between the two populations, can be developed for either of these methods.

OC Curves for the D2S Verification Method

In the D2S method, a test is performed on a single split sample to compare agency and contractor test results. If we assume both of these samples are from normally distributed subpopulations, then we can calculate the variance of the difference and use it to calculate two standard deviation, or approximately 95 percent, limits for the sample difference quantity. Suppose the agency subpopulation has a variance σ_A^2 and the contractor subpopulation has a variance σ_C^2. Since the variance of the difference in two random variables is the sum of the variances, the variance of the difference in an agency observation and a contractor observation is $\sigma_A^2 + \sigma_C^2$. The D2S limits are based on the test standard deviation provided. Let us call this test standard deviation σ_{test}. Under an assumption that $\sigma_A^2 = \sigma_C^2 = \sigma_{test}^2$, this variance of a difference becomes $2\sigma_{test}^2$. The D2S limits are set as two times the standard deviation (i.e., approximately 95 percent limits) of the test differences. This therefore sets that the D2S limits at $\pm 2\sqrt{2\sigma_{test}^2}$, which is $\pm 2\sqrt{2}\sigma_{test}$, or $\pm$ 2.8284 σ_{test}. Without loss of generality, we can assume $\sigma_{test} = 1$, along with assumption of a mean difference of 0, and use the standard normal distribution with region between –2.8284 and +2.8284 as acceptance region for the difference in an agency test result and a contractor test result. With these two limits fixed, we can calculate power of this decision–making process relative to various true differences in the underlying subpopulation means and/or various ratios of the true underlying subpopulation standard deviations.

These power values can conveniently be displayed as a three–dimensional surface. If we vary the mean difference along the first axis and the standard deviation ratio along a second axis, we can show power on the vertical axis. The agency subpopulation, the contractor subpopulation, or both, could have standard deviations smaller, about the same, or larger than the supplied σ_{test} value. Each of these cases is considered in the technical report for this project. (17) For simplicity, herein we will consider only the case where one of the two subpopulations has standard deviation equal to the supplied σ_{test}. Figure 50 shows the OC curves for this case. Power values are shown where the ratio of the larger of agency or contractor standard deviation to the smaller of agency or contractor standard deviation is varied over the values 0, 1, 2, 3, 4, and 5. The mean difference given along the horizontal axis (values 0, 1, 2, 3) represents the difference in agency and contractor subpopulation means expressed as multiples of σ_{test}.

As can be seen in the figure, even when the ratio of the contractor and agency standard deviations is 5 and the difference between the contractor and agency means is 3 times the value for σ_{test}, there is less than a 70 percent chance of detecting the difference based on the results from a single split sample.

As is the case with any method based on a sample of size one, the D2S method does not have much power to detect differences between the contractor and agency populations. The appeal of the D2S method lies in its simplicity rather than its power.

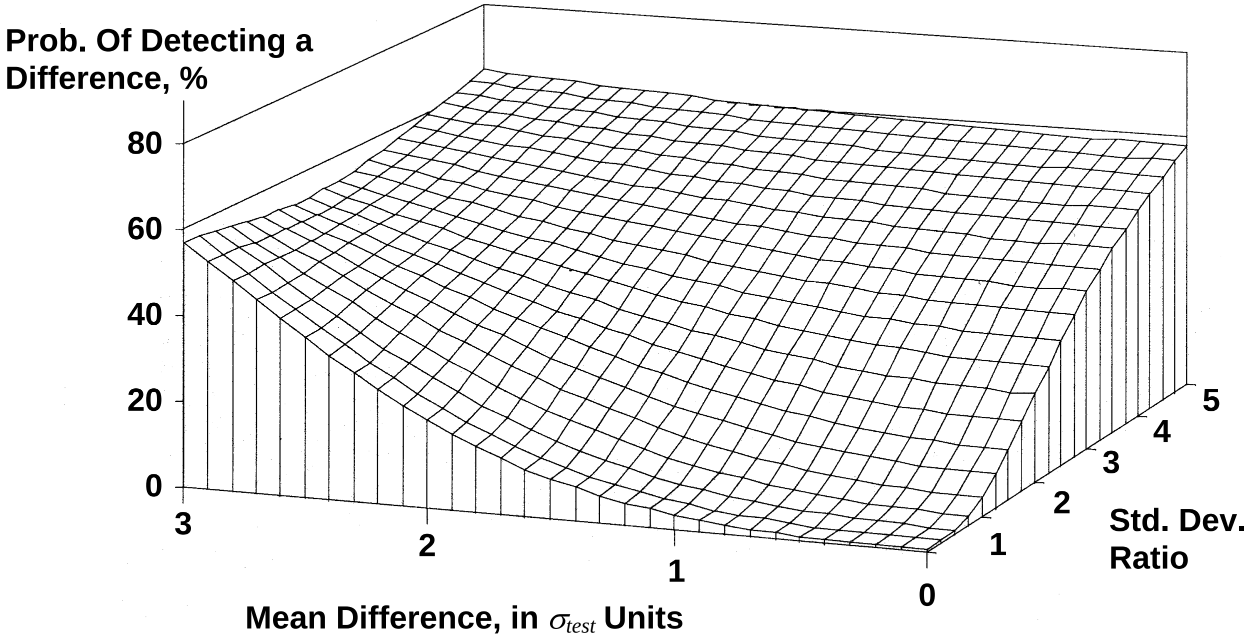


Figure 50. OC Surface for the D2S Test Method Verification Method (Assuming the smaller $\sigma = \sigma_{test}$)

OC Curves for the Paired *t*–test Method

As noted in chapter 5, for the case in which it is desirable to compare more than one pair of split sample test results, the *t*–test for paired measurements can be used. But the question arises, how many pairs of test results should be used? This is where an OC curve is helpful. The OC curve, for a given level of α, plots on the vertical axis either the probability of not detecting, β, or detecting, 1 – β, a difference between two populations. The standardized difference between the two population means is plotted on the horizontal axis.

For a *t*–test for paired measurements, the standardized difference, *d*, is measured as:

$$d = \frac{|\mu_c - \mu_a|}{\sigma_d} \tag{58}$$

where: $|\mu_c - \mu_a|$ = the true absolute difference between the mean of the contractor's test result population (which is unknown) and the mean of the agency's test result population (which is unknown).

σ_d = the standard deviation of the true population of signed differences between the paired tests (which is unknown).

The OC curves are developed for a given level of significance, α. It is evident from the OC curves that for any probability of not detecting a difference, β, (value on the vertical axis), the required *n* will increase as the difference, *d*, decreases (value on the horizontal axis). In some cases the desired β or *d* may require prohibitively large sample sizes. In that case a compromise must be made between the discriminating power desired, the cost of the amount of testing required, and the risk of claiming a difference when none exists.

OC curves for paired *t*–tests for α values of 0.05 and 0.01 appear in figures 51 and 52, respectively.

To use these OC curves the true standard deviation of the signed differences, σ_d, is assumed to be known, (or approximated based on published literature). After experience is gained with the process, σ_d can be more accurately defined and a better idea of the required number of tests determined.

Example 1. The number of pairs of split sample tests for verification of laboratory–compacted air voids using the Superpave Gyratory Compactor (SGC) is desired. The probability of **not** detecting a difference, β, is chosen as 20 percent or 0.20. (Some OC curves use 1 – β, known as the power of the test, on the vertical axis, but the only difference is the scale change, with 1 – β, in this case, being 80 percent). Assume that the absolute difference between μ_c and μ_a should not be greater than 1.25 percent, that the standard deviation using the SGC is 0.5 percent, and that α is selected as 0.01. This produces a *d* value of 1.25 percent/0.5 percent = 2.5. Reading this value on the horizontal axis and a β of 0.20 on the vertical axis in figure 52 shows that about 5 paired split–sample tests are necessary for the comparison.

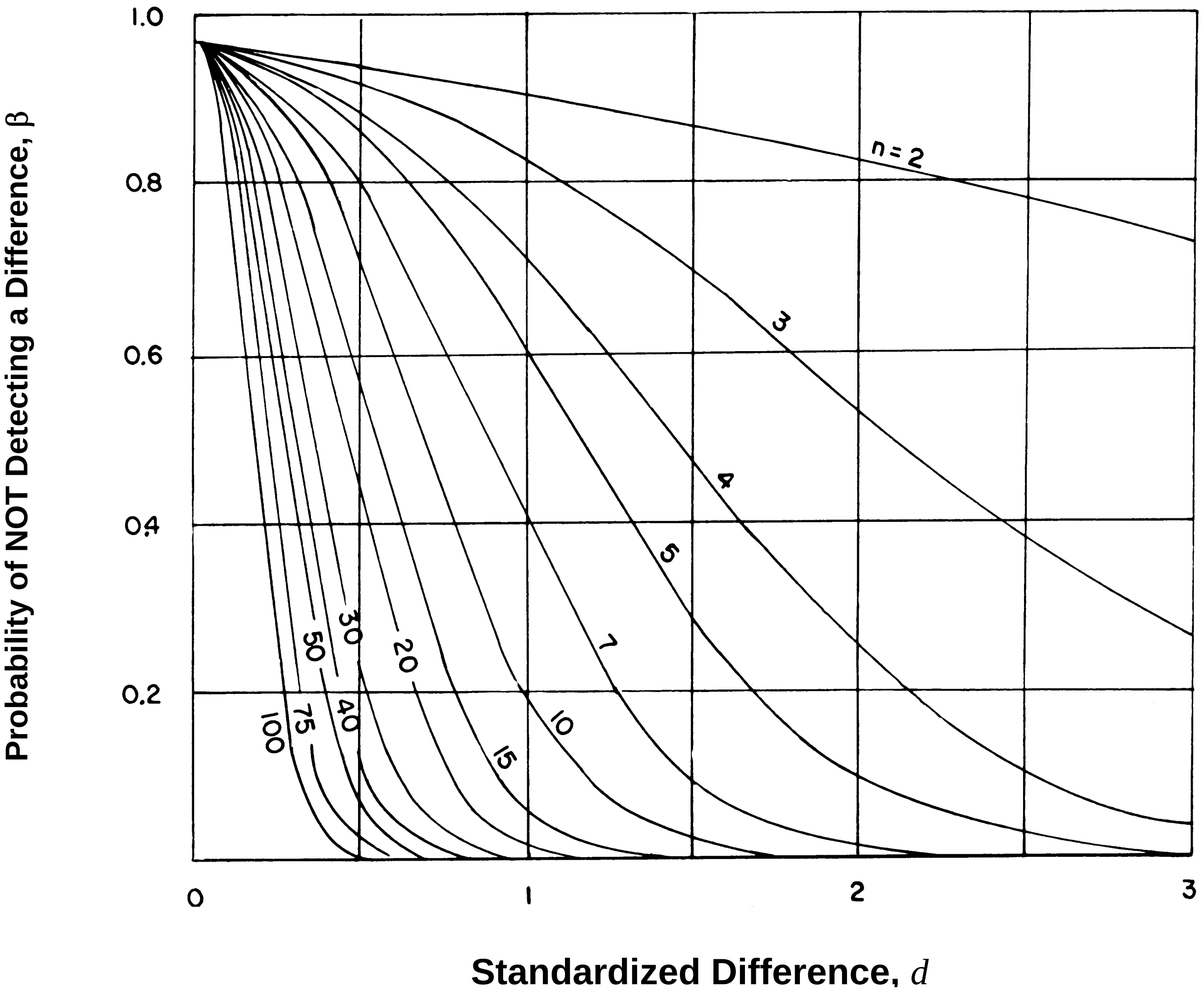


Figure 51. OC Curves for a Two–Sided *t*–Test (α = 0.05)

(Source: Experimental Statistics, by M. G. Natrella, National Bureau of Standards Handbook 91, 1963)

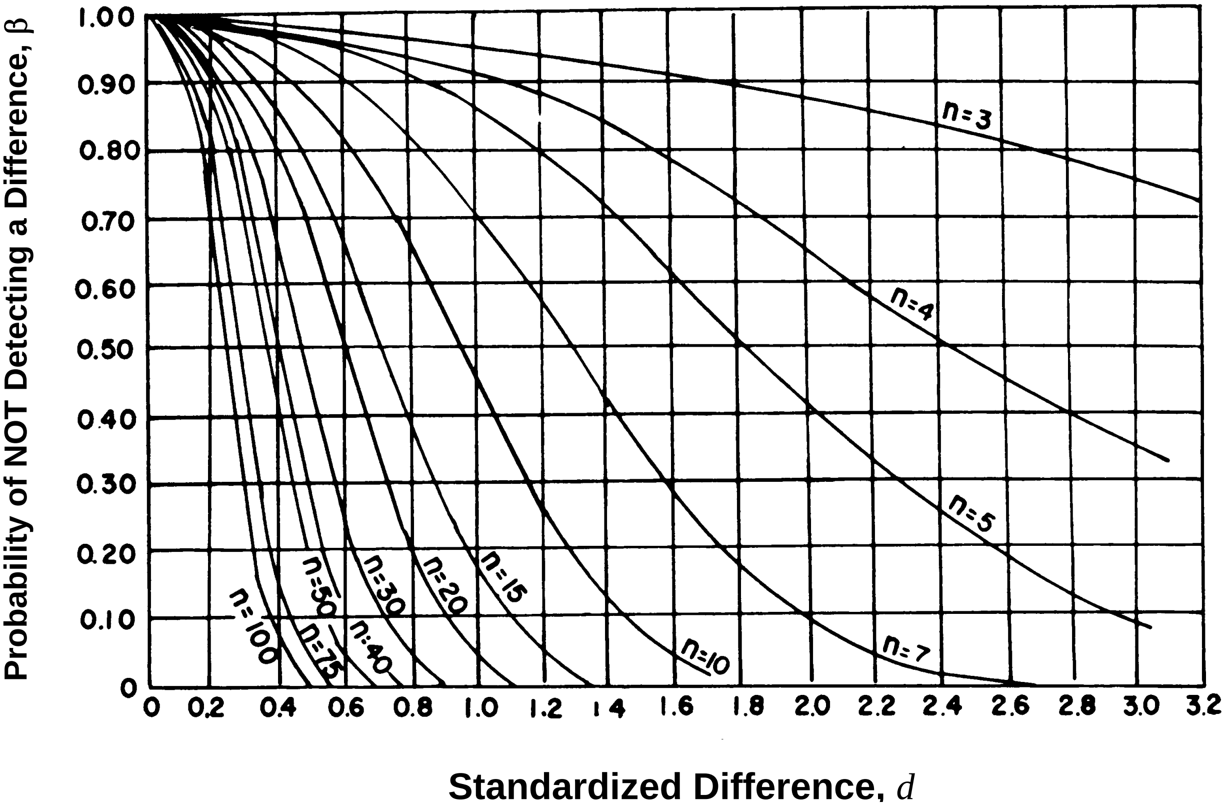


Figure 52. OC Curves for a Two–Sided t–Test ($\alpha = 0.01$)

(Source: Experimental Statistics, by M. G. Natrella,
National Bureau of Standards Handbook 91, 1963)

OC Curves for Process Verification

In chapter 5, two methods were considered for process verification using independently obtained samples: the *F*–test and *t*–test method, which compares the variances and means of sets of contractor and agency test results, and the single agency test method, which compares a single agency test result with 5 to 10 contractor test results. OC curves, which plot the probability of not detecting a difference, β, or detecting a difference, the power or 1 – β, versus the actual difference between the two populations, can be developed for either of these methods.

OC Curves for the *F*–test and *t*–test

One approach for comparing the contractor's test results with the agency's test results is to use the *F*–test and *t*–test comparisons of characteristics of the two data sets. To compare two populations that are assumed normally distributed, it is necessary to compare their means and their variabilities. An *F*–test is used to assess the size of the ratio of the variances, and a *t*–test is used to assess the degree of difference in the means. A question that needs to be answered is what power do these statistical tests have, when used with small to moderate size samples, to declare various differences in means and variances to be statistically significant differences. Some OC curves and examples of their use in power analysis follow.

***F*–test for Variances—Equal Sample Sizes.** Suppose we have two sets of measurements assumed to come from normally distributed populations and wish to conduct a test to see if they come from populations that have the same variances, i.e., $\sigma_{\mathrm{x}}^2 = \sigma_{\mathrm{y}}^2$. Further suppose we select a level of significance of α = .05, meaning we are allowing up to 5 percent chance of incorrectly deciding the variances are different when they really are the same. If we assume these two samples are

$$x_1, x_2, \ldots, x_{nx,} \text{ and } y_1 , y_2 , \ldots, y_{ny},$$

calculate sample variances s_x^2 and s_y^2, and construct

$$F = s_x^2 / s_y^2 , \qquad (59)$$

we would accept $H_o : \sigma_{\mathrm{x}}^2 = \sigma_{\mathrm{y}}^2$ for values of F in the interval

$$\left\lfloor F_{1-\alpha/2,nx-1,ny-1}, F_{\alpha/2,nx-1,ny-1} \right\rfloor.$$

For this two–sided or two–tailed test, figure 53 shows the probability we have accepted the two samples as coming from populations with the same variabilities. This probability is usually referred to as β, and the power of the test as 1 – β. Notice the horizontal axis is the quantity λ, where $\lambda = \sigma_x / \sigma_y$, the true standard deviation ratio. So for λ = 1, where the hypothesis of equal variance should certainly be accepted, it is accepted with probability 0.95, reduced from 1.0 only by the magnitude of our selected type I error risk, α. One major limiting factor for the use of figure 53 is the restriction that $n_x = n_y = n$.

Example 2. Suppose we have $n_x = 6$ contractor tests and $n_y = 6$ agency tests, conduct an $\alpha = 0.05$ level test, and accept that these two sets of tests represent populations with equal variances. What power did our test have to discern if the populations from which these two sets of tests came were really rather different in variabilities? Suppose the true population standard deviation of the contractor tests (σ_x) was twice as large as that of the agency tests (σ_y), giving $\lambda = 2$. If we enter figure 53 with $\lambda = 2$ and $n_x = n_y = 6$, we find that $\beta \approx 0.74$, or the power, $1 - \beta$, is about 0.26. This tells us that with samples of $n_x = 6$ and $n_y = 6$, we only have 26 percent chance of detecting a standard deviation ratio of 2 (and correspondingly a four–fold difference in variance) as being different.

Example 3. Suppose we are not at all comfortable with the power of 0.26 in Example 1, and so subsequently we increase the number of tests used. Suppose we now have $n_x = 20$ and $n_y = 20$. If we again consider $\lambda = 2$, we can determine from figure 53 the power of detecting these sets of tests as coming from populations with unequal variances to be over 0.8, approximately 82 percent to 83 percent. If we proceed to conduct our *F*–test with these two samples, and conclude the underlying variances are equal, we certainly feel much more comfortable with our conclusions.

Figure 54 gives the appropriate OC curves to use if we choose to conduct an $\alpha = 0.01$ level test. Again we see for equal variances σ_x^2 and σ_y^2, giving $\lambda = 1$, that $\beta = 0.99$, reduced from 1.0 only by the size of α.

***F*–test for Variances—Unequal Sample Sizes.** Up to now the discussions and OC curves presented have been limited to the case when the two sample sizes are equal. Calculation routines were developed for this project for calculation of power for this test for any combination of sample sizes n_x and n_y. There are obviously an infinite number of possible combinations for n_x and n_y. So, it is not possible to present OC curves for every possibility. However, three sets of tables are provided herein which provide a subset of power calculations using some sample sizes that are of potential interest for comparing contractor and agency samples. These power calculations are presented in table form since there are too many variables to present in a single chart, and the data can be presented in a more compact form in tables than in a long series of charts. Table 37 gives power values for all combinations of sample sizes from 3 to 10, with the ratio of the two subpopulation standard deviations being 1, 2, 3, 4, and 5. Table 38 gives power values for the same sample sizes, but with the standard deviation ratios being 0.0, 0.2, 0.4, 0.6, 0.8, and 1.0. Table 39 gives power values for all combinations for sample sizes of 5, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, and 100, with the standard deviation ratio being 1, 2, or 3. An example below illustrates the use of the first of these tables to reference power for a hypothetical test.

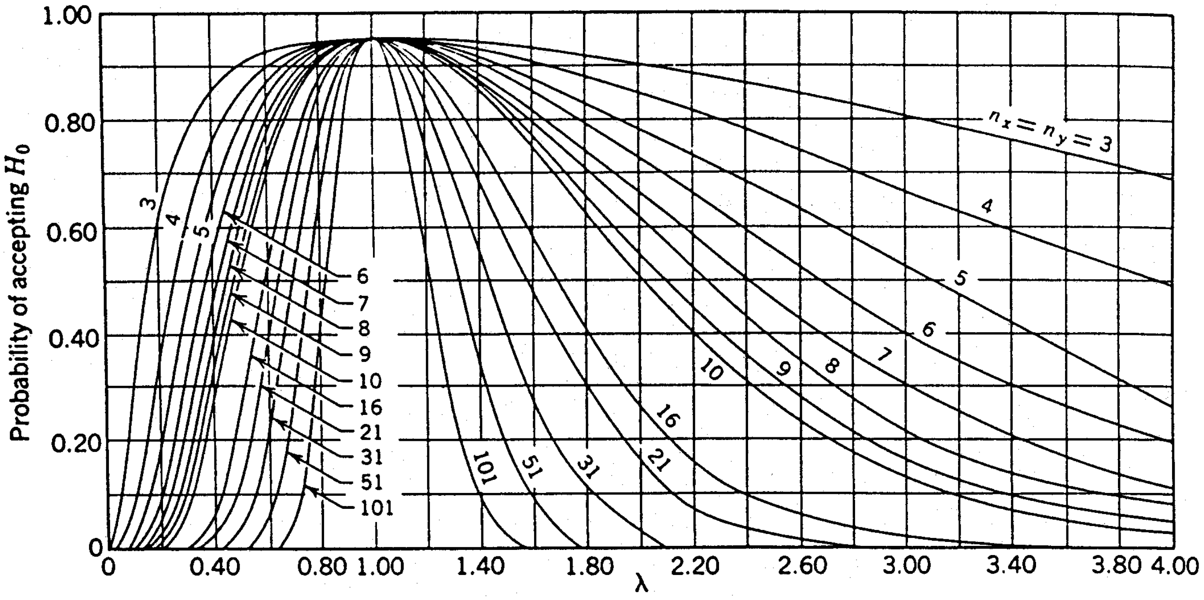


Figure 53. OC Curves for the Two–Sided *F*–Test for Level of Significance
α = 0.05 *(Source: Engineering Statistics by A. H. Bowker and G. J. Lieberman.)*

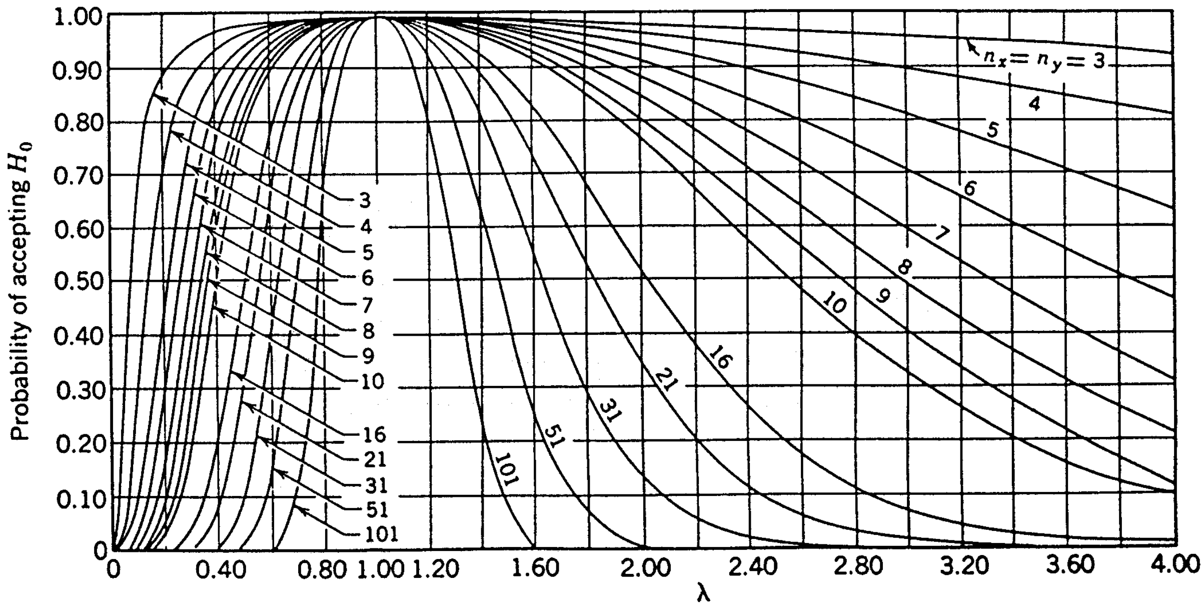


Figure 54. OC Curves for the Two–Sided *F*–Test for Level of Significance
α = 0.01 *(Source: Engineering Statistics by A. H. Bowker and G. J. Lieberman.)*

Table 37. *F*–test Power Values for n = 3 to 10 and *s*–ratio, λ = 1 to 5

λ	n_y	n_x	**Power**
1	3	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	4	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	5	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	6	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	7	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	8	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
1	9	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	10	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
2	3	3	0.09939
		4	0.09753
		5	0.09663
		6	0.09620
		7	0.09600
		8	0.09590
		9	0.09586
		10	0.09585
	4	3	0.14835
		4	0.15169
		5	0.15385
		6	0.15544
		7	0.15668
		8	0.15767
		9	0.15848
		10	0.15915
	5	3	0.19036
		4	0.20240
		5	0.21041
		6	0.21622
		7	0.22064
		8	0.22413
		9	0.22694
		10	0.22926
	6	3	0.22309
		4	0.24464
		5	0.25968
		6	0.27093
		7	0.27968
		8	0.28669
		9	0.29243
		10	0.29722
2	7	3	0.24820
		4	0.27854
		5	0.30055
		6	0.31744
		7	0.33086
		8	0.34179
		9	0.35087
		10	0.35853
	8	3	0.26768
		4	0.30567
		5	0.33401
		6	0.35619
		7	0.37410
		8	0.38888
		9	0.40129
		10	0.41187
	9	3	0.28308
		4	0.32758
		5	0.36144
		6	0.38837
		7	0.41036
		8	0.42869
		9	0.44421
		10	0.45752
	10	3	0.29549
		4	0.34549
		5	0.38414
		6	0.41521
		7	0.44081
		8	0.46230
		9	0.48060
		10	0.49639
3	3	3	0.19034
		4	0.19354
		5	0.19556
		6	0.19696
		7	0.19798
		8	0.19875
		9	0.19934
		10	0.19981
	4	3	0.31171
		4	0.33525
		5	0.35007
		6	0.36030
		7	0.36777
		8	0.37347
		9	0.37795
		10	0.38157

Table 37. *F*–test Power Values for n = 3 to 10 and *s*–ratio, λ = 1 to 5 *(cont.)*

λ	n_y	n_x	Power
3	5	3	0.39758
		4	0.44454
		5	0.47603
		6	0.49872
		7	0.51588
		8	0.52931
		9	0.54011
		10	0.54899
	6	3	0.45403
		4	0.51906
		5	0.56396
		6	0.59696
		7	0.62225
		8	0.64225
		9	0.65846
		10	0.67186
	7	3	0.49230
		4	0.57007
		5	0.62436
		6	0.66443
		7	0.69516
		8	0.71943
		9	0.73906
		10	0.75523
	8	3	0.51945
		4	0.60623
		5	0.66693
		6	0.71159
		7	0.74565
		8	0.77236
		9	0.79378
		10	0.81129
	9	3	0.53955
		4	0.63285
		5	0.69797
		6	0.74560
		7	0.78161
		8	0.80958
		9	0.83177
		10	0.84970
	10	3	0.55494
		4	0.65311
		5	0.72136
		6	0.77092
		7	0.80803
		8	0.83654
		9	0.85890
		10	0.87675
4	3	3	0.29251
		4	0.30367
		5	0.31010
		6	0.31427
		7	0.31717
		8	0.31930
		9	0.32093
		10	0.32222
	4	3	0.46558
		4	0.51179
		5	0.54104
		6	0.56126
		7	0.57608
		8	0.58742
		9	0.59637
		10	0.60363
	5	3	0.56455
		4	0.63665
		5	0.68356
		6	0.71649
		7	0.74084
		8	0.75955
		9	0.77437
		10	0.78638
	6	3	0.62143
		4	0.70759
		5	0.76314
		6	0.80150
		7	0.82932
		8	0.85027
		9	0.86652
		10	0.87943
	7	3	0.65697
		4	0.75074
		5	0.81002
		6	0.84993
		7	0.87808
		8	0.89866
		9	0.91416
		10	0.92613
	8	3	0.68090
		4	0.77901
		5	0.83976
		6	0.87961
		7	0.90692
		8	0.92628
		9	0.94042
		10	0.95100
4	9	3	0.69798
		4	0.79871
		5	0.85988
		6	0.89907
		7	0.92520
		8	0.94321
		9	0.95598
		10	0.96525
	10	3	0.71073
		4	0.81311
		5	0.87423
		6	0.91256
		7	0.93751
		8	0.95427
		9	0.96583
		10	0.97399
5	3	3	0.39165
		4	0.41270
		5	0.42481
		6	0.43266
		7	0.43815
		8	0.44219
		9	0.44530
		10	0.44776
	4	3	0.58713
		4	0.64932
		5	0.68814
		6	0.71467
		7	0.73394
		8	0.74858
		9	0.76007
		10	0.76932
	5	3	0.68068
		4	0.76196
		5	0.81171
		6	0.84479
		7	0.86811
		8	0.88527
		9	0.89836
		10	0.90860
	6	3	0.72975
		4	0.81790
		5	0.86956
		6	0.90223
		7	0.92409
		8	0.93936
		9	0.95041
		10	0.95864

Table 37. F–test Power Values for n = 3 to 10 and s–ratio, λ = 1 to 5 *(cont.)*

λ	n_y	n_x	Power
5	7	3	0.75893
		4	0.84940
		5	0.90024
		6	0.93086
		7	0.95030
		8	0.96318
		9	0.97201
		10	0.97824
	8	3	0.77800
		4	0.86909
		5	0.91845
		6	0.94695
		7	0.96423
		8	0.97513
		9	0.98225
		10	0.98704
	9	3	0.79133
		4	0.88238
		5	0.93024
		6	0.95690
		7	0.97244
		8	0.98184
		9	0.98772
		10	0.99150
	10	3	0.80115
		4	0.89188
		5	0.93838
		6	0.96351
		7	0.97767
		8	0.98594
		9	0.99092
		10	0.99400

Table 38. *F*–test Power Values for *n* = 3 to 10 and *s*–ratio, λ = 0.0 to 1.0

λ	n_y	n_x	Power
0.0	3	3	1.00000
		4	1.00000
		5	1.00000
		6	1.00000
		7	1.00000
		8	1.00000
		9	1.00000
		10	1.00000
	4	3	1.00000
		4	1.00000
		5	1.00000
		6	1.00000
		7	1.00000
		8	1.00000
		9	1.00000
		10	1.00000
	5	3	1.00000
		4	1.00000
		5	1.00000
		6	1.00000
		7	1.00000
		8	1.00000
		9	1.00000
		10	1.00000
	6	3	1.00000
		4	1.00000
		5	1.00000
		6	1.00000
		7	1.00000
		8	1.00000
		9	1.00000
		10	1.00000
	7	3	1.00000
		4	1.00000
		5	1.00000
		6	1.00000
		7	1.00000
		8	1.00000
		9	1.00000
		10	1.00000
	8	3	1.00000
		4	1.00000
		5	1.00000
		6	1.00000
		7	1.00000
		8	1.00000
		9	1.00000
		10	1.00000
	9	3	1.00000
		4	1.00000
		5	1.00000
		6	1.00000
		7	1.00000
		8	1.00000
		9	1.00000
		10	1.00000
	10	3	1.00000
		4	1.00000
		5	1.00000
		6	1.00000
		7	1.00000
		8	1.00000
		9	1.00000
		10	1.00000
0.2	3	3	0.39165
		4	0.58713
		5	0.68068
		6	0.72975
		7	0.75893
		8	0.77800
		9	0.79133
		10	0.80115
	4	3	0.41270
		4	0.64932
		5	0.76196
		6	0.81790
		7	0.84940
		8	0.86909
		9	0.88238
		10	0.89188
	5	3	0.42481
		4	0.68814
		5	0.81171
		6	0.86956
		7	0.90024
		8	0.91845
		9	0.93024
		10	0.93838
	6	3	0.43266
		4	0.71467
		5	0.84479
		6	0.90223
		7	0.93086
		8	0.94695
		9	0.95690
		10	0.96351
	7	3	0.43815
		4	0.73394
		5	0.86811
		6	0.92409
		7	0.95030
		8	0.96423
		9	0.97244
		10	0.97767
	8	3	0.44219
		4	0.74858
		5	0.88527
		6	0.93936
		7	0.96318
		8	0.97513
		9	0.98184
		10	0.98594
	9	3	0.44530
		4	0.76007
		5	0.89836
		6	0.95041
		7	0.97201
		8	0.98225
		9	0.98772
		10	0.99092
	10	3	0.44776
		4	0.76932
		5	0.90860
		6	0.95864
		7	0.97824
		8	0.98704
		9	0.99150
		10	0.99400
0.4	3	3	0.14221
		4	0.22806
		5	0.29564
		6	0.34398
		7	0.37868
		8	0.40429
		9	0.42380
		10	0.43906
	4	3	0.14250
		4	0.24034
		5	0.32488
		6	0.38884
		7	0.43614
		8	0.47159
		9	0.49879
		10	0.52015

Table 38. F–test Power Values for n = 3 to 10 and s–ratio, λ = 0.0 to 1.0 *(cont.)*

λ	n_y	n_x	Power
0.4	5	3	0.14291
		4	0.24808
		5	0.34448
		6	0.42028
		7	0.47749
		8	0.52079
		9	0.55411
		10	0.58029
	6	3	0.14332
		4	0.25345
		5	0.35863
		6	0.44371
		7	0.50889
		8	0.55851
		9	0.59674
		10	0.62671
	7	3	0.14369
		4	0.25739
		5	0.36934
		6	0.46187
		7	0.53357
		8	0.58837
		9	0.63057
		10	0.66355
	8	3	0.14399
		4	0.26041
		5	0.37772
		6	0.47638
		7	0.55351
		8	0.61261
		9	0.65804
		10	0.69341
	9	3	0.14424
		4	0.26278
		5	0.38447
		6	0.48825
		7	0.56996
		8	0.63266
		9	0.68076
		10	0.71805
	10	3	0.14445
		4	0.26470
		5	0.39001
		6	0.49813
		7	0.58375
		8	0.64952
		9	0.69984
		10	0.73868
0.6	3	3	0.07564
		4	0.10273
		5	0.12665
		6	0.14614
		7	0.16173
		8	0.17425
		9	0.18444
		10	0.19283
	4	3	0.07283
		4	0.10212
		5	0.13003
		6	0.15430
		7	0.17470
		8	0.19170
		9	0.20593
		10	0.21791
	5	3	0.07120
		4	0.10174
		5	0.13222
		6	0.15988
		7	0.18396
		8	0.20461
		9	0.22225
		10	0.23736
	6	3	0.07022
		4	0.10157
		5	0.13386
		6	0.16407
		7	0.19107
		8	0.21472
		9	0.23528
		10	0.25314
	7	3	0.06960
		4	0.10153
		5	0.13516
		6	0.16736
		7	0.19675
		8	0.22292
		9	0.24600
		10	0.26628
	8	3	0.06919
		4	0.10155
		5	0.13622
		6	0.17003
		7	0.20139
		8	0.22972
		9	0.25499
		10	0.27741
0.6	9	3	0.06891
		4	0.10161
		5	0.13711
		6	0.17223
		7	0.20526
		8	0.23545
		9	0.26265
		10	0.28698
	10	3	0.06870
		4	0.10168
		5	0.13786
		6	0.17409
		7	0.20854
		8	0.24035
		9	0.26925
		10	0.29529
0.8	3	3	0.05467
		4	0.06163
		5	0.06758
		6	0.07248
		7	0.07649
		8	0.07980
		9	0.08255
		10	0.08487
	4	3	0.05202
		4	0.05929
		5	0.06587
		6	0.07156
		7	0.07642
		8	0.08057
		9	0.08412
		10	0.08719
	5	3	0.05017
		4	0.05755
		5	0.06448
		6	0.07067
		7	0.07612
		8	0.08090
		9	0.08508
		10	0.08875
	6	3	0.04883
		4	0.05626
		5	0.06340
		6	0.06995
		7	0.07584
		8	0.08109
		9	0.08577
		10	0.08994

Table 38. *F*–test Power Values for *n* = 3 to 10 and *s*–ratio, λ = 0.0 to 1.0 *(cont.)*

λ	n_y	n_x	Power
0.8	7	3	0.04785
		4	0.05529
		5	0.06258
		6	0.06938
		7	0.07560
		8	0.08124
		9	0.08633
		10	0.09092
	8	3	0.04709
		4	0.05453
		5	0.06193
		6	0.06893
		7	0.07541
		8	0.08136
		9	0.08680
		10	0.09175
	9	3	0.04650
		4	0.05393
		5	0.06141
		6	0.06856
		7	0.07527
		8	0.08148
		9	0.08721
		10	0.09248
	10	3	0.04603
		4	0.05345
		5	0.06099
		6	0.06827
		7	0.07516
		8	0.08159
		9	0.08757
		10	0.09312
1.0	3	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	4	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
1.0	5	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	6	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	7	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	8	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	9	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000
	10	3	0.05000
		4	0.05000
		5	0.05000
		6	0.05000
		7	0.05000
		8	0.05000
		9	0.05000
		10	0.05000

Table 39. *F*–test Power Values for *n* = 5 to 100 and *s*–ratio, λ = 1 to 3

λ	n_y	n_x	Power
1	5	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
	10	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
	15	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
1	20	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
	25	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
	30	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
1	40	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
	50	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
	60	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05

Table 39. F–test Power Values for n = 5 to 100 and s–ratio, λ = 1 to 3 *(cont.)*

λ	n_y	n_x	**Power**
1	70	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
	80	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
	90	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05

λ	n_y	n_x	**Power**
1	100	5	0.05
		10	0.05
		15	0.05
		20	0.05
		25	0.05
		30	0.05
		40	0.05
		50	0.05
		60	0.05
		70	0.05
		80	0.05
		90	0.05
		100	0.05
2	5	5	0.21041
		10	0.22926
		15	0.23658
		20	0.24043
		25	0.24281
		30	0.24442
		40	0.24646
		50	0.24770
		60	0.24853
		70	0.24913
		80	0.24958
		90	0.24993
		100	0.25022
	10	5	0.38414
		10	0.49639
		15	0.55109
		20	0.58353
		25	0.60501
		30	0.62027
		40	0.64053
		50	0.65336
		60	0.66221
		70	0.66869
		80	0.67363
		90	0.67753
		100	0.68068

λ	n_y	n_x	**Power**
2	15	5	0.45487
		10	0.62152
		15	0.70573
		20	0.75560
		25	0.78820
		30	0.81099
		40	0.84054
		50	0.85870
		60	0.87092
		70	0.87969
		80	0.88626
		90	0.89137
		100	0.89545
	20	5	0.49087
		10	0.68548
		15	0.78230
		20	0.83747
		25	0.87192
		30	0.89495
		40	0.92304
		50	0.93906
		60	0.94918
		70	0.95606
		80	0.96099
		90	0.96468
		100	0.96753
	25	5	0.51241
		10	0.72299
		15	0.82516
		20	0.88085
		25	0.91389
		30	0.93485
		40	0.95864
		50	0.97099
		60	0.97817
		70	0.98272
		80	0.98578
		90	0.98795
		100	0.98955

Table 39. *F*–test Power Values for *n* = 5 to 100 and *s*–ratio, λ = 1 to 3 *(cont.)*

λ	n_y	n_x	Power
2	30	5	0.52669
		10	0.74730
		15	0.85174
		20	0.90637
		25	0.93725
		30	0.95585
		40	0.97551
		50	0.98476
		60	0.98968
		70	0.99256
		80	0.99436
		90	0.99556
		100	0.99639
	40	5	0.54439
		10	0.77664
		15	0.88220
		20	0.93379
		25	0.96067
		30	0.97548
		40	0.98924
		50	0.99462
		60	0.99702
		70	0.99821
		80	0.99886
		90	0.99923
		100	0.99945
	50	5	0.55491
		10	0.79358
		15	0.89881
		20	0.94770
		25	0.97160
		30	0.98387
		40	0.99414
		50	0.99757
		60	0.99888
		70	0.99943
		80	0.99969
		90	0.99982
		100	0.99989
2	60	5	0.56187
		10	0.80456
		15	0.90914
		20	0.95588
		25	0.97764
		30	0.98820
		40	0.99632
		50	0.99869
		60	0.99948
		70	0.99977
		80	0.99989
		90	0.99995
		100	0.99997
	70	5	0.56683
		10	0.81224
		15	0.91614
		20	0.96120
		25	0.98137
		30	0.99073
		40	0.99745
		50	0.99921
		60	0.99972
		70	0.99989
		80	0.99996
		90	0.99998
		100	0.99999
	80	5	0.57053
		10	0.81791
		15	0.92118
		20	0.96490
		25	0.98387
		30	0.99235
		40	0.99810
		50	0.99947
		60	0.99984
		70	0.99994
		80	0.99998
		90	0.99999
		100	1.00000
2	90	5	0.57339
		10	0.82226
		15	0.92497
		20	0.96762
		25	0.98564
		30	0.99345
		40	0.99851
		50	0.99962
		60	0.99989
		70	0.99997
		80	0.99999
		90	1.00000
		100	1.00000
	100	5	0.57568
		10	0.82571
		15	0.92793
		20	0.96968
		25	0.98696
		30	0.99425
		40	0.99879
		50	0.99972
		60	0.99993
		70	0.99998
		80	0.99999
		90	1.00000
		100	1.00000
3	5	5	0.47603
		10	0.54899
		15	0.57700
		20	0.59187
		25	0.60108
		30	0.60736
		40	0.61537
		50	0.62026
		60	0.62355
		70	0.62593
		80	0.62772
		90	0.62911
		100	0.63024

Table 39. *F*–test Power Values for *n* = 5 to 100 and *s*–ratio, λ = 1 to 3 *(cont.)*

λ	n_y	n_x	Power
3	10	5	0.72136
		10	0.87675
		15	0.92836
		20	0.95158
		25	0.96404
		30	0.97154
		40	0.97985
		50	0.98420
		60	0.98681
		70	0.98853
		80	0.98973
		90	0.99062
		100	0.99130
	15	5	0.78336
		10	0.93786
		15	0.97640
		20	0.98918
		25	0.99431
		30	0.99669
		40	0.99860
		50	0.99928
		60	0.99957
		70	0.99972
		80	0.99980
		90	0.99985
		100	0.99988
	20	5	0.80975
		10	0.95808
		15	0.98816
		20	0.99597
		25	0.99841
		30	0.99930
		40	0.99982
		50	0.99994
		60	0.99998
		70	0.99999
		80	0.99999
		90	1.00000
		100	1.00000
3	25	5	0.82417
		10	0.96743
		15	0.99254
		20	0.99797
		25	0.99936
		30	0.99977
		40	0.99996
		50	0.99999
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000
	30	5	0.83321
		10	0.97267
		15	0.99463
		20	0.99877
		25	0.99968
		30	0.99990
		40	0.99999
		50	1.00000
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000
	40	5	0.84390
		10	0.97822
		15	0.99654
		20	0.99938
		25	0.99987
		30	0.99997
		40	1.00000
		50	1.00000
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000
3	50	5	0.84999
		10	0.98107
		15	0.99738
		20	0.99960
		25	0.99993
		30	0.99999
		40	1.00000
		50	1.00000
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000
	60	5	0.85393
		10	0.98279
		15	0.99783
		20	0.99971
		25	0.99996
		30	0.99999
		40	1.00000
		50	1.00000
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000
	70	5	0.85668
		10	0.98394
		15	0.99812
		20	0.99976
		25	0.99997
		30	1.00000
		40	1.00000
		50	1.00000
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000

Table 39. *F*–test Power Values for *n* = 5 to 100 and *s*–ratio, λ = 1 to 3 *(cont.)*

λ	n_y	n_x	Power
3	80	5	0.85871
		10	0.98476
		15	0.99831
		20	0.99980
		25	0.99998
		30	1.00000
		40	1.00000
		50	1.00000
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000
	90	5	0.86026
		10	0.98537
		15	0.99844
		20	0.99983
		25	0.99998
		30	1.00000
		40	1.00000
		50	1.00000
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000
	100	5	0.86150
		10	0.98584
		15	0.99855
		20	0.99985
		25	0.99998
		30	1.00000
		40	1.00000
		50	1.00000
		60	1.00000
		70	1.00000
		80	1.00000
		90	1.00000
		100	1.00000

Example 4. Suppose we have $n_x = 10$ contractor tests and $n_y = 6$ agency tests, conduct an $\alpha = 0.05$ level test, and accept that these two tests represent populations with equal variances. What power did our test have to discern if the populations from which these two sets of tests came were really rather different in variabilities? Suppose the true population standard deviation of the contractor's test population (σ_x) was twice as large as that of the agency's test population (σ_y), giving a standard deviation ratio value, $\lambda = 2$. If we enter table 37 with $\lambda = 2$, $n_x = 10$, and $n_y = 6$, we find the power to be 0.41521. This tells us that with samples of $n_x = 10$ and $n_y = 6$, we have slightly less than a 42 percent chance of detecting a standard deviation ratio of 2 (and correspondingly a four–fold difference in variances) as being different.

***t*–test for Means.** Suppose we have two sets of measurements, assumed to be from normally distributed populations, and wish to conduct a two–sided or two–tailed test to see if these populations have equal means, i.e., $\mu_x = \mu_y$. Suppose we assume these two samples are from populations with unknown, but equal, variances. If these two samples are $x_1, x_2, \ldots, x_{nx}$ with sample mean $\overline{X}$ and sample variance s_x^2, and $y_1, y_2, \ldots, y_{ny}$ with sample mean $\overline{Y}$ and sample variance s_y^2, we can calculate

$$t = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{s_x^2(n_x - 1) + s_y^2(n_y - 1)}{n_x + n_y - 2}} \times \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}} \tag{60}$$

and accept H_o: $\mu_x = \mu_y$ for values of t in the interval $[-t_{\alpha/2, n_x + n_y - 2}, t_{\alpha/2, n_x + n_y - 2}]$.

For this test, figure 51 or 52, depending upon the α value, shows the probability we have accepted the two samples as coming from populations with the same means. The horizontal axis scale is

$$d = \frac{|\mu_x - \mu_y|}{\sigma} \tag{61}$$

where $\sigma = \sigma_x = \sigma_y$ is the true common population standard deviation. We access the OC curves in figure 51 or 52 with a value for d of d^* and a value for n of n' where

$$n' = n_x + n_y - 1$$

and (62)

$$d^* = \frac{d}{\sqrt{n'}} \times \sqrt{\frac{n_x \times n_y}{n_x + n_y}}$$

Example 5. Suppose we have $n_x = 8$ contractor tests and $n_y = 8$ agency tests, conduct an $\alpha = 0.05$ level test, and accept that these two sets of tests represent populations with equal means. What power did our test really have to discern if the populations from which these two sets of tests came had different means? Suppose we consider a difference in these population means of 2 or more standard deviations as a noteworthy difference that we would like to detect with high probability. This would indicate that we are interested in $d = 2$. Calculating

$$n' = n_x + n_y - 1 = 8 + 8 - 1 = 15 \text{ and}$$

$$d^* = \frac{d}{\sqrt{n'}} \times \sqrt{\frac{n_x \times n_y}{n_x + n_y}} = \frac{2}{\sqrt{15}} \times \sqrt{\frac{8 \times 8}{8+8}} = 1.0328 \tag{63}$$

we find from figure 51 that $\beta \approx 0.05$ so that our power of detecting a mean difference of 2 or more σ would be approximately 95 percent.

Example 6. Suppose we consider an application where we still have a total of 16 tests, but with $n_x = 12$ contractor tests and $n_y = 4$ agency tests. Suppose that we are again interested in *t*–test performance in detecting a means difference of 2 standard deviations. Again

$$n' = n_x + n_y - 1 = 12 + 4 - 1 = 15,$$

but now (64)

$$d^* = \frac{d}{\sqrt{n'}} \times \sqrt{\frac{n_x \times n_y}{n_x + n_y}} = \frac{2}{\sqrt{15}} \times \sqrt{\frac{12 \times 4}{12+4}} = 0.8944$$

We find from figure 51 that $\beta \approx 0.12$ indicating a power of approximately 88 percent of detecting a mean difference of 2 or more standard deviations.

Figure 52 gives the appropriate OC curves for our use in conducting an $\alpha = 0.01$ level test on means. This figure is accessed in the same manner as described above for figure 51.

OC Curves for the Single Agency Test Method

This procedure involves comparing the mean of 5 to 10 contractor tests with a single agency test result. The two are considered to be similar if the agency test is within an allowable interval on either side of the mean of the contractor's test results. The allowable interval is determined by multiplying the sample range of the contractor's test results by a factor that depends on the number of contractor test results. The equations for computing the allowable intervals are shown in table 40.

This comparison method is adapted from an approach for calculating the confidence interval for estimating a population mean. A confidence interval for a population mean is calculated about a sample mean and defines an interval within which there is a given percent confidence that the true population mean falls. When the variability of the population is unknown, a *t*–distribution, rather than a normal distribution, is used to calculate the confidence interval for the population mean. The *t*–distribution is what is used to establish the critical values for the *t*–statistic that is used in the *t*–test procedure that was presented above.

When calculating a confidence interval for the population mean, the *t*–statistic, which is similar in general concept to the *Z*–statistic of a normal distribution, is used. The *t*–statistic depends upon the degrees of freedom, defined as $n - 1$ where n is the number of values used to obtain the sample mean. The confidence interval is defined by:

$$\overline{X} \pm \left(t \times \frac{s}{\sqrt{n}} \right) \tag{65}$$

The value of t depends upon the number of degrees of freedom and the level of significance chosen for the confidence interval. For example, for a 98 percent confidence interval, the value of t would be the value such that 98 percent of a *t*–distribution with $n - 1$ degrees of freedom fell within the mean and $\pm t$ standard deviations.

The single agency test approach uses this 98 percent confidence interval to approximate the interval within which a single test result should fall if sampled from a population with mean and standard deviation equal to the sample mean and standard deviation of the contractor's test results. For simplicity, the sample range, R, instead of the sample standard deviation, is used to estimate the population standard deviation. The population standard deviation can be estimated by dividing the sample range by a factor known as d_2. Therefore, $R \div d_2$ is taken as an estimate of the population standard deviation.

The approach assumes that the population mean is equal to the sample mean of the contractor's tests and that the population standard deviation is equal to the contractor's sample range divided by d_2. The interval within which the single agency test result must fall is defined by the interval within which 98 percent of the single test results should fall. The 98 percent confidence interval is calculated based on the *t*–statistic.

To arrive at the factors in the table for determining the interval around the contractor's test mean within which the agency test must fall, the *t*–statistic for a 98 percent confidence interval and $n - 1$ degrees of freedom is multiplied by $(R \div d_2)$. Since it is a two–sided confidence interval, a 98 percent confidence interval corresponds to the $\pm$ *t*–statistic, $t_{.99}$, above or below which there is only 1 percent of the *t*–distribution. The values necessary to develop the interval factors for this comparison method are shown in table 40.

Table 40. Derivation of the Single Agency Test Method Allowable Intervals

Sample Size, *n*	**Degrees of Freedom, *n* – 1**	***t*–statistic for which there is a 1% chance of being exceeded, $t_{.99}$**	**d_2**	**Interval $I = \overline{X}_n \pm \left(\frac{t_{.99}}{d_2}\right) R$**
10	9	2.821	3.078	$I = \overline{X}_{10} \pm 0.91R$
9	8	2.896	2.970	$I = \overline{X}_9 \pm 0.97R$
8	7	2.998	2.847	$I = \overline{X}_8 \pm 1.05R$
7	6	3.143	2.704	$I = \overline{X}_7 \pm 1.17R$
6	5	3.365	2.534	$I = \overline{X}_6 \pm 1.33R$
5	4	3.747	2.326	$I = \overline{X}_5 \pm 1.61R$

To illustrate the lack of power that this method has to discern differences between populations, the computer program ONETEST was developed as part of FHWA Demonstration Project 89.(18) The ONETEST program assumes that the two sets of data have the same standard deviation value (an assumption that is part of the single test comparison method), and designates in standard deviation units the distance between the true means of the two datasets. The program then determines the probability of detecting the difference for various actual differences between the population means.

ONETEST was used to generate 6,000 comparisons for each of a number of different scenarios, i.e., comparing a single test result to samples of size 10, 9, 8, 7, 6, and 5. In each case, the two populations were assumed to have the same standard deviation, and the difference between the means of the two populations, stated in standard deviation units, $\Delta = (\mu_1 - \mu_2)/\sigma$, varied from 0.0 to 3.0 in increments of 0.5. The results from this analysis are plotted as an OC curve in figure 55.

As can be seen in the OC curve in figure 55, even when the difference between population means was three standard deviations, the percentage of the time this procedure was able to determine a difference in populations ranged from only 58 percent for a sample size of 10 to 34 percent for a sample size of 5.

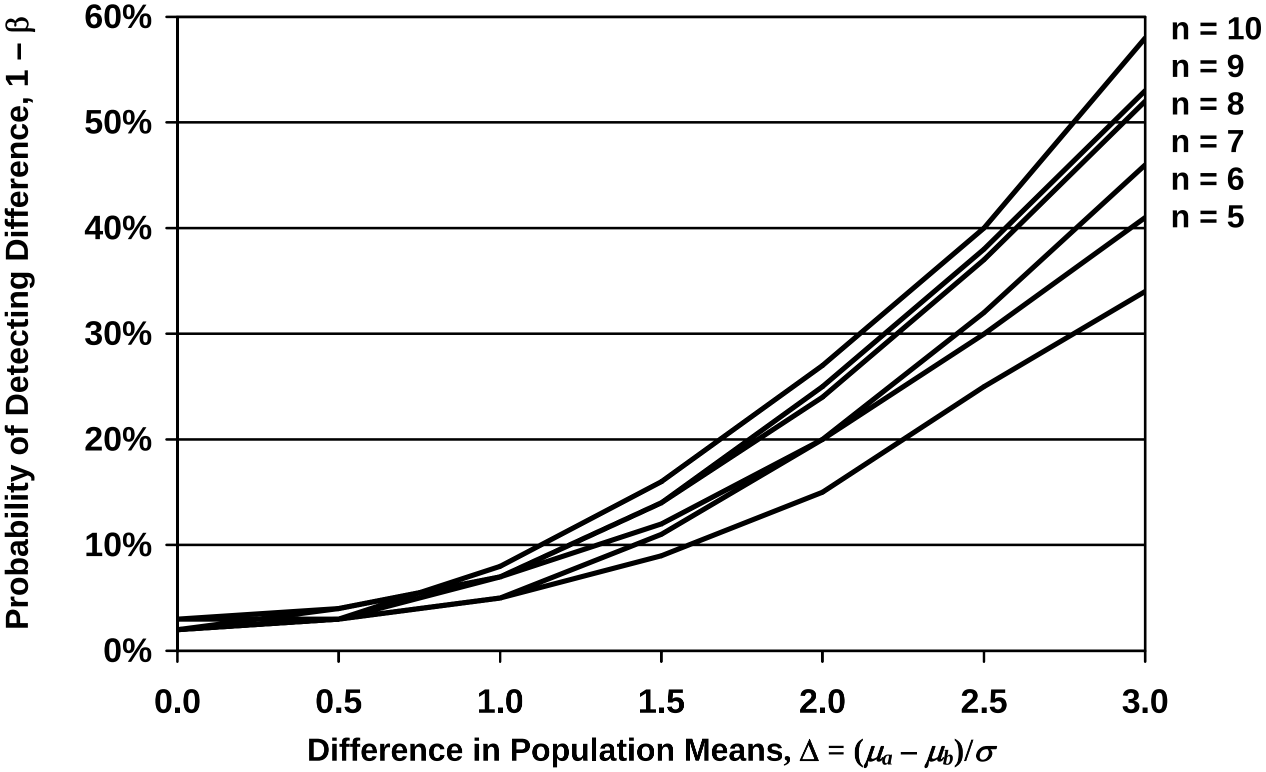


Figure 55. OC Curves for the Single Agency Test Method

APPENDIX H

Simple Methods for Evaluating the Normality of a Set of Data

Visual Observation of Histograms

A large group of measurements and test results cannot provide any useful information until they are organized in preparation for analysis. Until the data are organized into a form that is intelligible and understandable, they are just a collection of numbers. The human mind cannot easily comprehend a large series of separate facts or numbers.

A frequency histogram for a set of observations is a diagram that shows the frequency of occurrence of the values of the variable in ordered classes. Each group of observations is called a class. The frequency for any class is the number of observations with measurements falling within that class, while the relative frequency for any class is the frequency for that class divided by the total number of observations (data values).

Individual rectangles whose heights are proportional to the frequencies in each class are erected on the horizontal axis. The base of each rectangle is set equal to the class intervals. If the class intervals are equal in width, the area of an individual rectangle represents the number of observations within the class, while the total area under the figure represents the total number of data values. Figure 56 shows an example frequency histogram for a set of data.

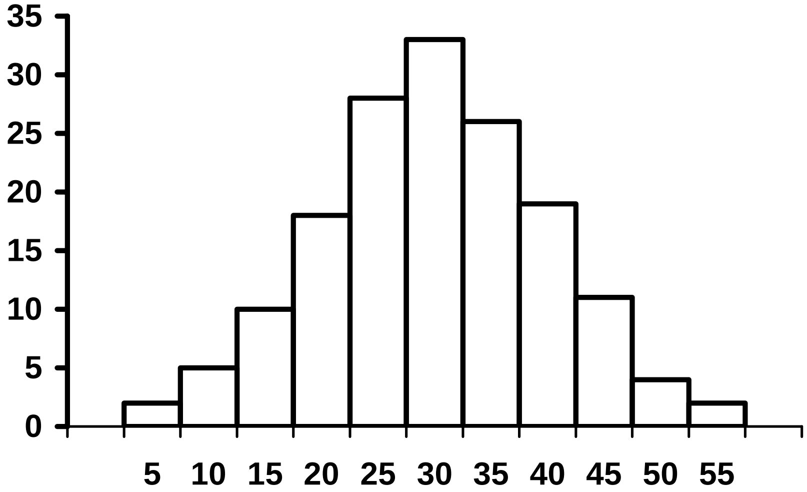


Figure 56. Example Frequency Histogram

Sometimes, a frequency histogram may be all that is needed to validate an assumption of a normally distributed data set. Figure 56, for example, appears to be approximately normally distributed. On the other hand, a frequency histogram can also show that a set of data is not normally distributed. This is the case for the histogram shown in figure 57, where the data are skewed to the left.

As mentioned above, it is reasonable to assume that most construction materials are approximately normally distributed. A study done for FHWA on both HMAC and PCC projects examined the occasions that skewed results occurred. (27) For PCC, few material properties of the 21 measured on 3 projects were significantly skewed. For the HMAC projects, out of 52 material properties measured on 3 projects, 6 (11.5 percent) had values that indicated a significant skewness. Five of these were from gradation results. One potential source of skewness is the presence of a physical barrier, such as occurs with the top size sieve. Since a gradation cannot exceed 100 percent, when the mean approaches this limiting value, the distribution will typically appear to be skewed. (See figure 57.)

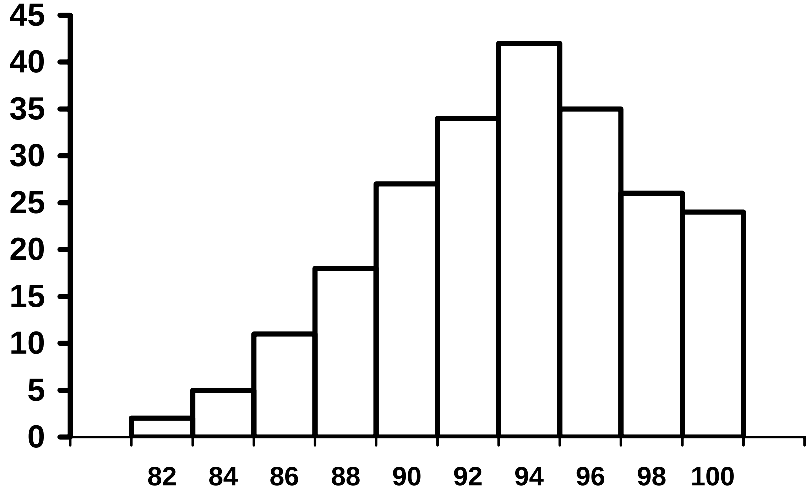


Figure 57. Example of a Skewed Frequency Histogram

Normal Probability Paper

Another visual approach to assessing the normality of a set of data is to plot the data on normal probability paper. Normal probability paper is graph paper for which the scales are established such that a normal distribution will plot as a straight line. Specifically, the cumulative distribution of a normal distribution will plot as a straight line on normal probability paper. The data can be plotted on the normal probability paper either as grouped data, such as that from a histogram plot, or as individual data points. A sample of normal probability paper is shown in figure 58.

If the cumulative frequency data plot as reasonably close to a straight line on normal probability paper, then it would be assumed that a normal distribution can be used as a reasonable approximation for the data set.

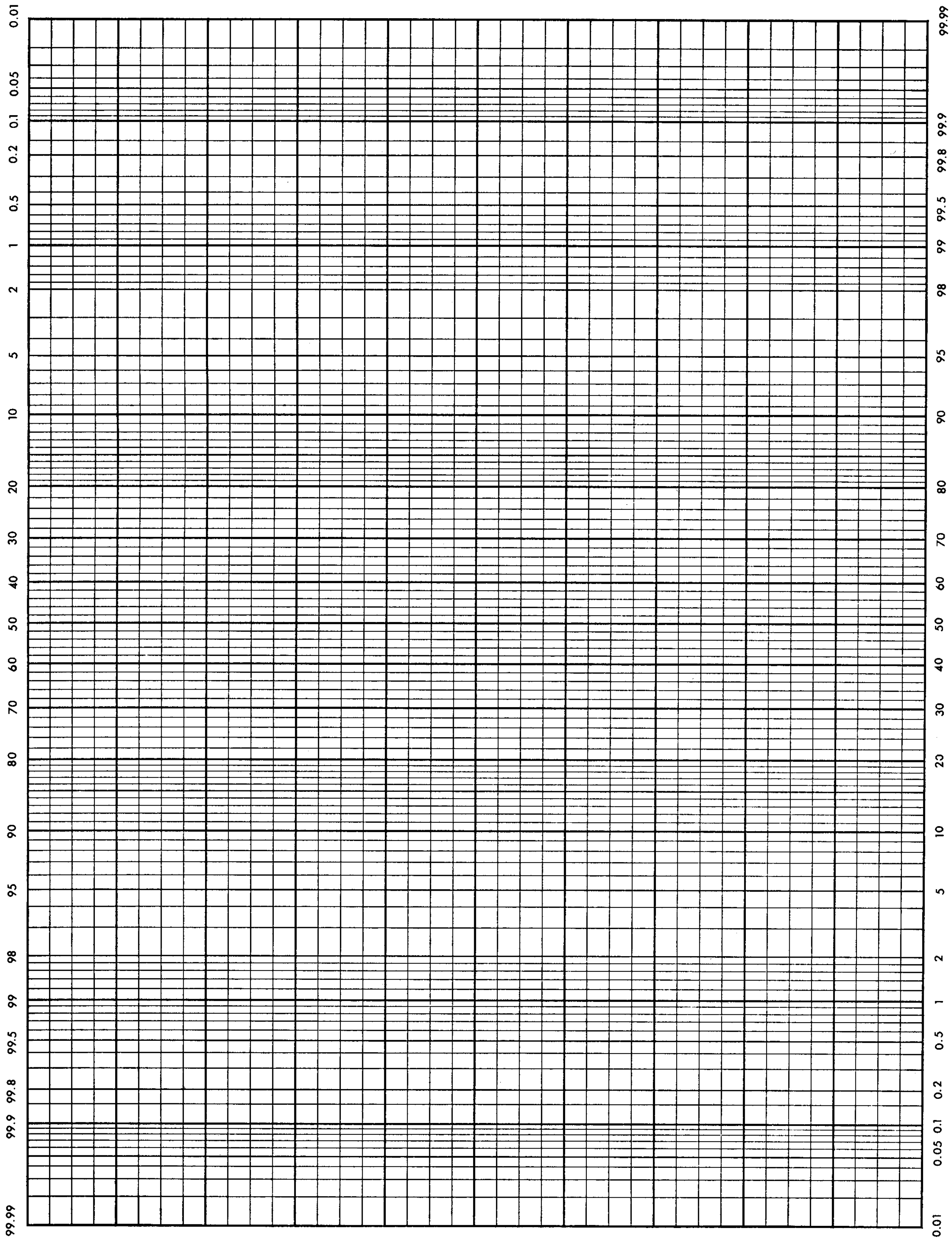


Figure 58. Normal Probability Paper

Grouped Data Example

Table 41 presents a set of 25 data points that have been grouped into 5 classes. The last column shows the cumulative relative frequency values. For example, the table shows that 14/25 = 0.56, or 56 percent, of the data points are less than or equal to 100.

Table 41. Example Cumulative Frequency Table for Grouped Data

Class Limits	Class Frequency	Cum. Frequency	Cum. Relative Frequency
86 – 93	4	4	4/25 = 16%
93 – 100	10	14	14/25 = 56%
100 – 107	9	23	23/25 = 92%
107 – 114	1	24	24/25 = 96%
114 – 121	1	25	25/25 = 100%

Figure 59 shows the results when the cumulative relatively frequency values from table 41 are plotted on normal probability paper against the respective upper class limits for each class. It should be noted that the cumulative relative frequency corresponding to the 121 upper class limit has been plotted at the 99.99 percent level. Using this approximation, the “best straight line” has been drawn through the points by ignoring the data point that represents the upper class limit of 114. Since only 5 plotted points are available, it is difficult to determine whether or not this point should be neglected. This situation can be improved, however, by either plotting all of the individual data points or by dividing the data into more classes.

Ungrouped Data Example

The plot of the grouped data in figure 59 did not clearly indicate whether or not the data could be approximated by a normal distribution. Plotting the individual data values in their ungrouped format can help to rectify this situation. The calculations for determining the cumulative relative frequencies for the individual data values are shown in table 42.

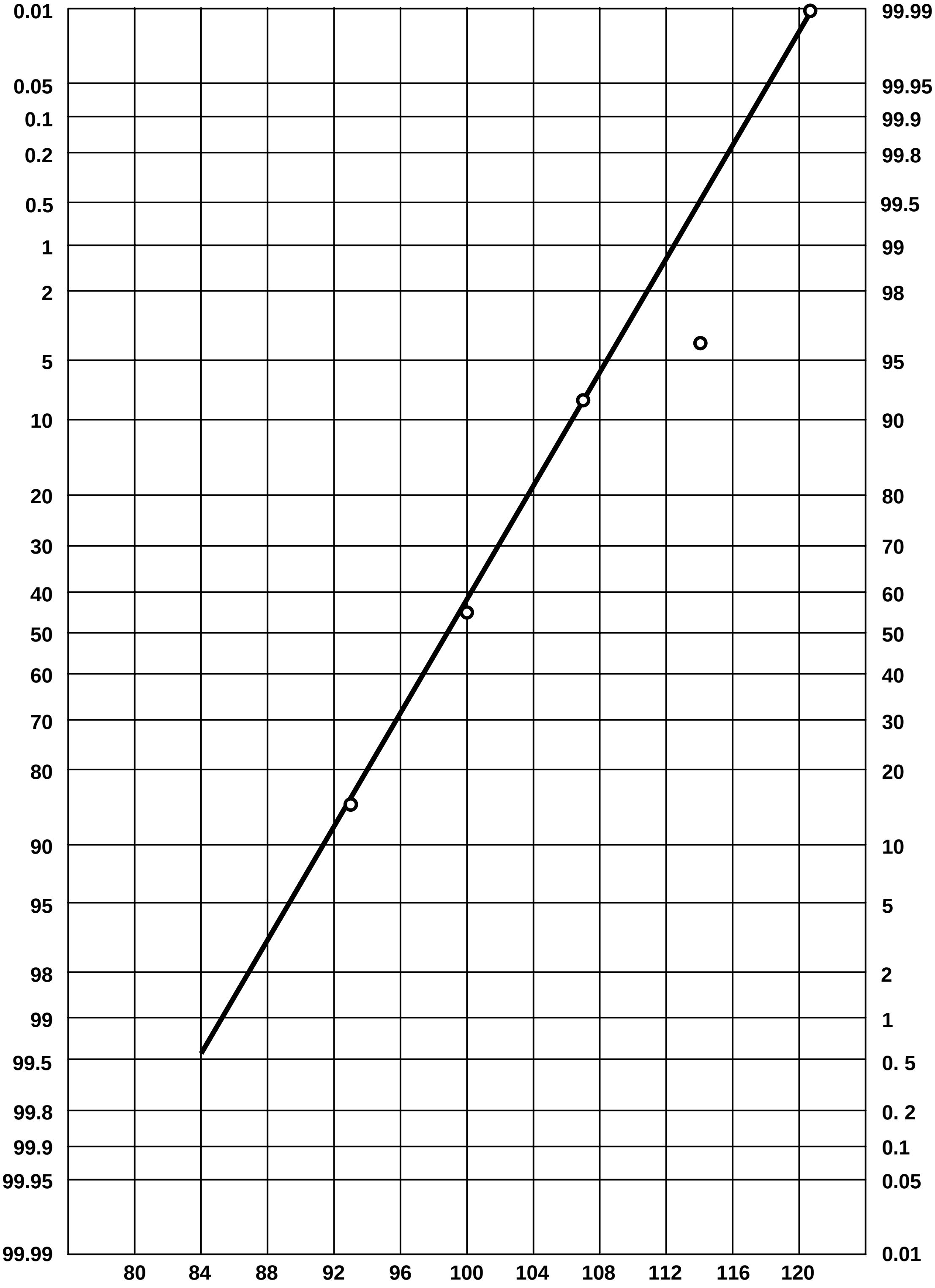


Figure 59. Normal Probability Plot for Grouped Data Example

Table 42. Example Cumulative Frequency Table for Ungrouped Data

Data Value	Cumulative Frequency	Cumulative Relative Frequency
87	1	1/26 = 4%
89	2	2/26 = 8%
91	3	3/26 = 12%
92	4	4/26 = 15%
93	5	5/26 = 19%
94	6	6/26 = 23%
95	7	7/26 = 27%
96	—	—
96	9	9/26 = 35%
97	10	10/26 = 38%
98	—	—
98	12	12/26 = 46%
99	13	13/26 = 50%
100	14	14/26 = 54%
101	15	15/26 = 58%
102	16	16/26 = 62%
103	—	—
103	18	18/26 = 69%
104	—	—
104	20	20/26 = 77%
105	21	21/26 = 81%
106	22	22/26 = 85%
107	23	23/26 = 88%
112	24	24/26 = 92%
120	25	25/26 = 96%

It should be noted that a different method of calculation, one that is sometimes felt to provide a more realistic representation, has been used. Based on a small set of data, it would not be valid to create the impression that none (i.e., 0 percent) of the data are below 87 or that all (i.e., 100 percent) of the data are below 120. The common practice is therefore to add one more value to the number of observations (i.e., in our case 25 + 1 = 26), and then to compute the cumulative relative frequencies on that basis. As a result, there will be no 100 percent value. Typically, the plot of cumulative frequency distribution would be dotted below the lowest value and above the highest value.

The results from table 42 are plotted in figure 60. It appears, based on these data, that the assumption of normality is reasonable. A straight line fits the data reasonably well with the exception of the values in the upper tail (112 and 120 values).

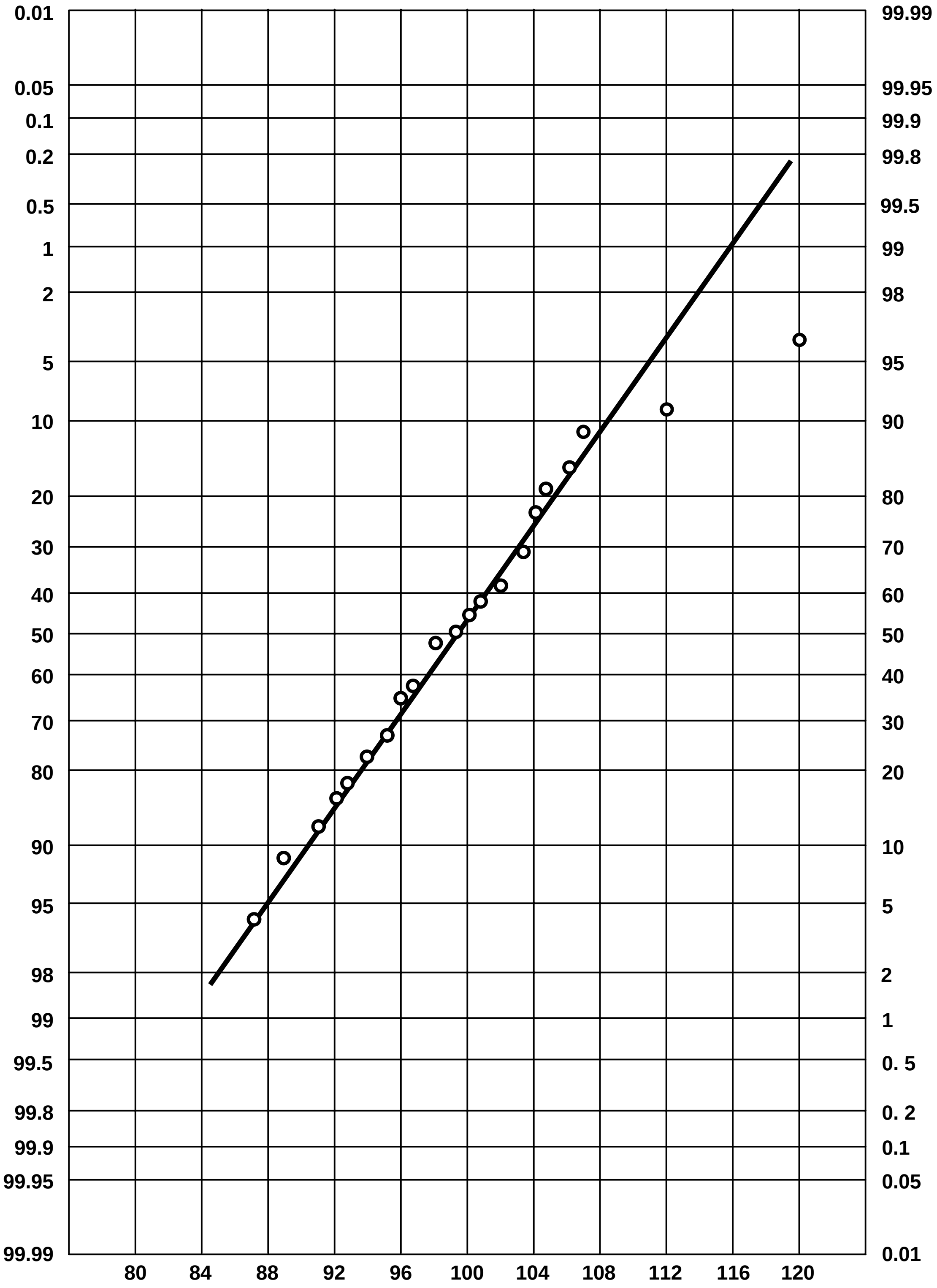


Figure 60. Normal Probability Plot for Ungrouped Data Example

Matching Moments

The methods discussed so far are graphical in nature, and require some degree of subjectivity when deciding whether the shape of the histogram is reasonably normal or whether the data plot as a straight line on normal probability paper. There are other, more quantitative, methods available for considering whether or not a set of data is normally distributed. In many cases, the graphical methods will be sufficient since it is only necessary that the data are approximately normal.

One method for evaluating whether or not a set of data is reasonably normal is sometimes called the method of matching moments. A normality test based on moment measures was proposed in the 1920's. (28) The first and second central moments are often used in statistical analyses to calculate the mean and the variance. The third and fourth central moments are less frequently used, and represent a measure of symmetry and kurtosis, respectively. While tables of critical values for conducting normality GOF tests using moments have been developed, (29) it is not anticipated that a highway agency would choose this method for GOF testing. A Chi Square or Kolmogorov–Smirnov test would more likely be chosen if a formal GOF test procedure were desired.

The third moment is a measure of asymmetry and is called skewness. The fourth moment is a measure of kurtosis, or the "peakedness" of the distribution. Kurtosis looks at how much of the total distribution lies in the tails of the distribution. Skewness and kurtosis coefficients have been developed such that the normal distribution has skewness and kurtosis values of 0. Therefore, a highway agency could calculate the skewness and kurtosis for a data set to see how close the values for the data set are to 0.

Skewness.

The TRB glossary (2) includes the following definition:

- ***Skewness**—a measure of the symmetry of a distribution. When the distribution has a greater tendency to tail to the right, it is said to have positive skewness. When the distribution has a greater tendency to tail to the left, it is said to have negative skewness. For the normal distribution (as well as for any other symmetrical distribution), the skewness coefficient equals 0.*

  Population skewness coefficient: $$\gamma_1 = \sum (X_i - \mu)^3 / 2n\sigma^3 \qquad (66)$$

  Sample skewness coefficient: $$g_1 = n\sum (X_i - \bar{X})^3 / [s^3(n-1)(n-2)] \qquad (67)$$

Therefore, the skewness characterizes the degree of asymmetry of a distribution around its mean. Positive skewness indicates a distribution with a long tail extending toward more positive values. Negative skewness indicates a distribution with a long tail extending toward more negative values. The calculations for skewness can be unwieldy, particularly for large data sets. It is therefore recommended that a computer program, such as the Microsoft® Excel spreadsheet program, be used for calculating the skewness. The equation for sample skewness used in Excel, which is algebraically the same as the equation used in the TRB glossary, (2) is as follows:

$$g_1 = \frac{n}{(n-1)(n-2)}\left[\sum_{i=1}^{i=n}\left(\frac{X_i - \overline{X}}{s}\right)^3\right] \tag{68}$$

where: g_1 = skewness.
n = total number of data values.
X_i = individual data values.
$\overline{X}$ = mean of the set of data values.
s = standard deviation of the set of data values.

Kurtosis

The TRB glossary (2) includes the following definition:

- ***Kurtosis**—a measure of the shape of a distribution. For the normal distribution, the kurtosis coefficient equals 0. A positive kurtosis coefficient indicates that the distribution has longer tails than the normal distribution, while a negative coefficient indicates that the distribution has shorter tails.*

  Population kurtosis coefficient:

  $$\gamma_2 = [\sum(X_1 - \mu)^4 / n\sigma^4] - 3 \tag{69}$$

  Sample kurtosis coefficient:

  $$g_2 = [n(n+1)\sum(X_1 - \overline{X})^4 / s^4(n-1)(n-2)(n-3)] - [3(n-1)^2 / (n-2)(n-3)] \tag{70}$$

The above definition is a little confusing. The definition refers to distributions with tails longer or shorter than those of a normal distribution. In theory, the normal distribution runs from minus infinity to plus infinity. It is, therefore, not possible to have tails “longer” than the normal distribution.

A better explanation is that kurtosis characterizes the relative amount of the distribution that is in the tails of the distribution, i.e., the “weight” of the tails of the distribution. A normal distribution has a kurtosis coefficient equal to zero. Negative kurtosis indicates distributions where a larger proportion of the values are towards the extremes, i.e., relatively “fat” or “heavy” tails compared with a normal distribution. Positive kurtosis, on the other hand, indicates distributions where the values are bunched up near the mean, i.e., relatively “thin” or “light” tails compared with a normal distribution. The calculations for kurtosis can be unwieldy, particularly for large data sets. It is therefore recommended that a computer program, such as the Microsoft® Excel spreadsheet program, be used for calculating the kurtosis. The equation for sample kurtosis used in Excel, which is algebraically the same as the equation used in the TRB glossary, (2) is as follows:

$$g_2 = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \left[\sum_{i=1}^{i=n} \left(\frac{X_i - \overline{X}}{s} \right)^4 \right] \right\} - \left(\frac{3(n-1)^2}{(n-2)(n-3)} \right) \tag{71}$$

where: g_2 = kurtosis.
n = total number of data values.
X_i = individual data values.
$\overline{X}$ = mean of the set of data values.
s = standard deviation of the set of data values.

GOF Tests

The previously mentioned methods for evaluating normality all require some degree of subjectivity on the part of the evaluator. GOF tests exist that allow for a normality decision to be made with a given level of significance, α. These GOF tests provide a more objective and rigorous method for evaluating normality. It is recommended that a highway agency first plot a histogram, consider the skewness and kurtosis, and if necessary plot their data on normal probability paper to make the decision regarding the normality of a set of data.

If a decision is not obvious from the above measures, then a formal GOF test should be conducted. GOF tests are not considered in this manual. The most common are the Chi Square test and the Kolmogorov–Smirnov (K–S) test. These are both explained and described in detail in numerous statistical texts.

APPENDIX I

Derivation of Equation for Cost of Premature Pavement Failure

Equation to Be Derived

$$PAYADJ = \frac{C(R^D - R^E)}{(1 - R^O)} \qquad (72)$$

where: $PAYADJ$ = appropriate payment adjustment for new pavement or overlay (same units as C).

C = present total cost of resurfacing. (typical value = \$23.92/m^2).

D = design life of pavement or initial overlay (typically 20 years for new pavement, 10 years for overlay).

E = expected life of pavement or overlay (variable).

O = expected life of successive overlays (typically 10 years).

R = (1 + INF) / (1 + INT).

INF = long–term annual inflation rate in decimal form (typically 0.04).

INT = long–term annual interest rate in decimal form (typically 0.08).

Basic LCC Equations

The standard LCC equation is given by equation 73:

$$NPV = \frac{C}{(1 + DISC)^n} \qquad (73)$$

where: NPV = the net present value.
C = the present cost of the action in question.
$DISC$ = the annual discount rate in decimal form.
n = the number of years in the future that the action will take place.

Equation 73 is a very close approximation of what actually happens, and a slightly more accurate expression can be derived. If C is the cost of an action today, then its cost n years in the future is given by $C_n = C\ (1 + INF)^n$ in which INF is the annual inflation rate expressed as a decimal. The net present value of this future action is $NPV = C_n\ /\ (1 + INT)^n$ in which INT is the annual interest rate in decimal form. Combining these two equations produces $NPV = C\ (1 + INF)^n\ /\ (1 + INT)^n$ which, by defining $R = (1 + INF)\ /\ (1 + INT)$, becomes

$$NPV = C\ R^n \qquad (74)$$

To demonstrate that equation 74 produces almost exactly the same result as the standard LCC equation given by equation 73, typical long–term values of $INF = 0.04$ and $INT = 0.08$ will be used. Since the discount rate is conventionally treated as the difference between interest and inflation, $DISC = 0.08 - 0.04 = 0.04$ for this example. The following comparison can then be made for an action $n = 10$ years in the future:

equation 73: $$NPV = \frac{C}{(1+DISC)^n} = \frac{C}{1.04^{10}} = 0.676C \quad \leftarrow$$

equation 74: $$R = (1 + INF) / (1 + INT) = 1.04 / 1.08 = 0.963$$

$$NPV = C\,(0.963)^{10} = 0.686\,C \quad \leftarrow$$

It is seen that the results are virtually identical, thus explaining the common practice in LCC analysis of defining the discount rate as the difference between the interest and inflation rates. Because equation 74 is slightly more accurate, it is used in the derivation that follows. However, it is first used to demonstrate that equation 72 is valid.

Demonstration of Validity of Equation 72

One way to check the validity of a derived expression is to create a hypothetical test case that can readily be calculated by hand, and then compare it to the result obtained with the derived equation. A convenient test case for equation 72 is the case in which a new pavement, designed to last $D = 20$ years, fails at exactly $E = 10$ years at which time it will be resurfaced. Historical data have shown that an overlay typically lasts about 10 years. Therefore, this unplanned–for overlay will restore the pavement to its originally intended design life of 20 years, after which it will continue to receive overlays at approximate 10–year intervals. In this special case, the only added expense beyond what has already been anticipated is the extra overlay in the 10th year. Using a more accurate value of $R = (1 + INF) / (1 + INT) = 1.04 / 1.08 = 0.96296$, the net present cost of this additional overlay can be determined by equation 72 (the derived equation) as

$$PAYADJ = -23.92\,(0.96296^{20} - 0.96296^{10}) / (1 - 0.96296^{10}) = -\$16.40/\text{m}^2$$

or, by equation 74 (the more accurate version of the standard LCC analysis equation) as

$$PAYADJ = -23.92\,(0.96296)^{10} = -\$16.40/\text{m}^2.$$

The exact agreement between the more accurate version of the standard LCC analysis approach and the derived equation provides convincing evidence that the derived equation will also provide the correct answer in other situations for which the standard LCC analysis method is not convenient, or would require an assumption about residual value. However, the actual proof is contained in the derivation that follows.

Derivation of Equation 72

To derive equation 72, it is convenient to first write the expression for the net debit resulting from the rescheduling of any particular future overlay. For example, if a 20–year design pavement fails after 15 years, the two primary LCC analysis components associated with the rescheduling of the 20th–year overlay can be expressed as follows using equation 74 and the terms previously defined:

$$DEBIT\,(1) \;=\; C\,R^{15} \qquad (75)$$

$$CREDIT(1) \;=\; C\,R^{20} \qquad (76)$$

The net debit due to the rescheduling of the first overlay is then the difference of these two values or $NETDEBIT(1) = C\,(R^{20} - R^{15})$. The series of debits due to the successive rescheduling of the future overlays, which are expected to occur at approximate 10–year intervals, is as follows:

$$NETDEBIT(1) \;=\; C\,(R^{20} - R^{15}) \qquad (77)$$

$$NETDEBIT(2) \;=\; C\,(R^{30} - R^{25}) \qquad (78)$$

$$NETDEBIT(3) \;=\; C\,(R^{40} - R^{35}) \qquad (79)$$

$$NETDEBIT(4) \;=\; \text{etc.}$$

The total value in net–present–value terms is the sum of all the individual debits that, by segregating positive and negative terms, can be written as follows:

$$NPV \;=\; C\,[(R^{20} + R^{30} + R^{40} + \ldots) - (R^{15} + R^{25} + R^{35} + \ldots)] \qquad (80)$$

At this point it is convenient to write the equation in general terms:

$$NPV \;=\; C\,[(R^{D} + R^{D+O} + R^{D+2O} + \ldots) - (R^{E} + R^{E+O} + R^{E+2O} + \ldots)] \qquad (81)$$

This can be factored and rearranged as follows:

$$NPV \;=\; C\,[R^{D}(1 + R^{O} + R^{2O} + \ldots) - R^{E}(1 + R^{O} + R^{2O} + \ldots)] \qquad (82)$$

$$NPV \;=\; C\,(R^{D} - R^{E})\,(1 + R^{O} + R^{2O} + \ldots) \qquad (83)$$

The last term in parentheses in equation 83 is recognizable as a geometric progression. Because long–term inflation rates are always less than long–term interest rates, the ratio *R* will always be less than unity, thus causing the geometric progression to have a finite sum, as given by equation 84. (30)

$$1 + R^{O} + R^{2O} + \ldots \;=\; 1\,/\,(1 - R^{O}) \qquad (84)$$

Substituting this back into equation 83 completes the derivation, as shown in equation 85. In equation 72, the term *PAYADJ* is used in place of *NPV* because this equation applies to situations in which the change in expected life is estimated from construction characteristics that are under the contractor's control, thus providing the basis for rational payment–adjustment schedules. It should be noted that, in addition to justifying payment reductions for deficient quality, it appropriately awards incentive payments when superior quality extends the expected life of the pavement beyond the intended design life.

$$NPV = C\,(R^D - R^E) / (1 - R^O) \qquad (85)$$

Other Factors

Two factors have been ignored in this derivation. One is relatively minor, but the other could have a substantial impact on the true costs. Routine annual maintenance costs have not been included because they are relatively small in comparison to construction costs, and may change very little when the overlays are rescheduled. User delay costs, however, may not only be large, they may be magnified when premature failure results in additional, unplanned–for overlays. Although they are difficult to quantify, it is quite apparent that delay costs can only add to the true cost of premature failure, suggesting that payment adjustments even larger than those produced by equation 72 are justified.

Typical Results

Using typical long–term values for interest and inflation of 0.08 and 0.04, respectively, the ratio *R* is found to be 1.04/1.08 = 0.963. Using this in either of equations 72 or 85, and assuming a present total cost of resurfacing of \$23.97/m^2, a design life for resurfacing of 10 years, and a typical overlay life of 10 years, the values in table 43 are obtained. It is seen from this table that the cost of premature failure can be substantial, terminating in this case at the initial cost of resurfacing of \$23.92 \$/m^2 for zero expected life.

Table 43. Payment Adjustment as a Function of Expected Life

Expected Life, years	Appropriate Payment Adjustment, \$/m^2
12	+3.79
10 (Design)	0.00
8	–4.09
6	–8.50
4	–13.27
2	–18.40
0	–23.92

APPENDIX J

Polynomial Method to Model Performance Relationships

Background

Performance–related specifications require mathematical models to link construction quality to expected life and, ultimately, to value expressed in the form of payment schedules. Although ongoing research efforts continue to advance the state of the art, the type of data needed to develop accurate and precise models may not become available for years. In the interim, present engineering and mathematical knowledge can be used to create rational and practical models that can perform effectively until better models are available. Examples are presented to illustrate how both analytical data and survey data can be used to develop realistic performance models useful for the development of payment schedules for QA specifications. The issue of the proper method to combine the effects of multiple deficiencies is also addressed.

Development of a Rational Rejection Provision

The RQL is essentially defined as a severely deficient level of quality at which the agency reserves the option to require removal and replacement of the construction item. In the HMAC pavement specification for one agency, the RQL for both air voids and thickness had been defined in terms of percent defective as PD ≥ 75. In other words, if either air voids or thickness exhibited this level of quality, the lot could be declared rejectable. However, this leads to the inconsistency shown in table 44.

Table 44. An Inconsistent RQL Provision

Case	**Quality Level**		**Rejectable ?**
	Air Voids	**Thickness**	
1	PD = 0 (Excellent)	PD = 75 (RQL)	Yes
2	PD = 75 (RQL)	PD = 0 (Excellent)	Yes
3	PD = 74 (Almost RQL)	PD = 74 (Almost RQL)	No

Clearly, case 3 is far worse than the other two but, under the existing system, it would not trigger the RQL provision whereas the first two cases would. Defining the RQL in a more appropriate way rectified this inconsistency. Intuitively, if both air voids and thickness reach some intermediate value less than PD = 75, say $PD_{VOIDS} = PD_{THICK} = 50$, for example, then that might logically be just as detrimental as $PD_{VOIDS} = 0$, $PD_{THICK} = 75$ or $PD_{VOIDS} = 75$, $PD_{THICK} = 0$. To illustrate how a more suitable RQL provision can be developed, suppose the agency has determined that the conditions listed in table 45 are all likely to severely shorten the life of the pavement and, therefore, are appropriate RQL points.

Table 45. Example of Data Used to Develop Joint RQL Provision for Two Quality Measures

Quality Level		Rejectable ?
Air Voids	**Thickness**	
PD = 75	PD = 10	Yes
PD = 10	PD = 75	Yes
PD = 50	PD = 50	Yes

By plotting these three points on a graph, as illustrated by model #1 in figure 61, it can be seen that the model must be in the form of a curve that is concave–downward.

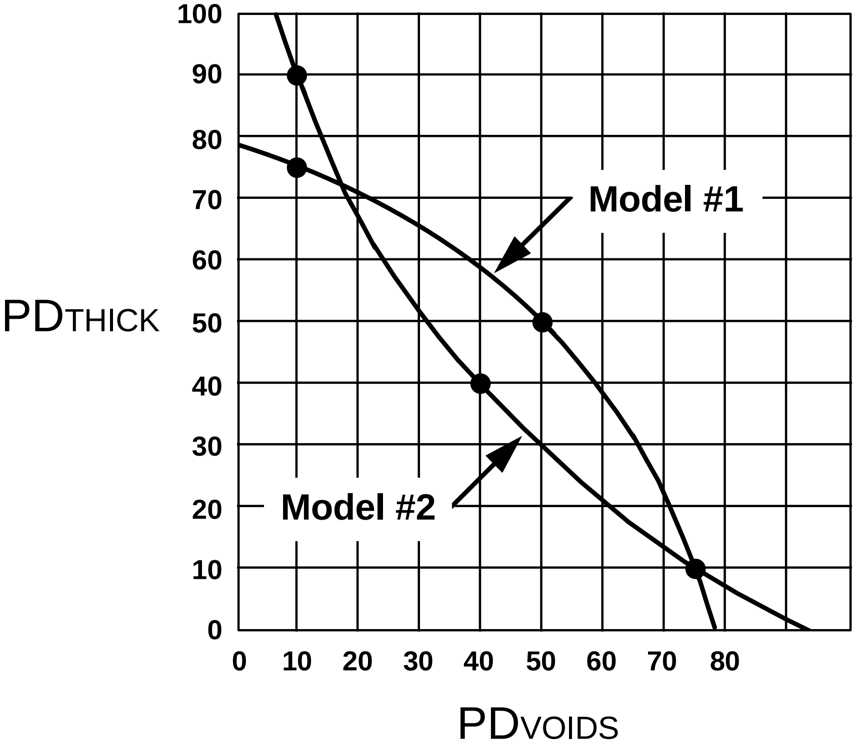


Figure 61. Graph of Two Possible Mathematical Models for an RQL Provision

Since the purpose of this model is to account for the combined effect of air voids and thickness, a simple way to accomplish this is to include the cross–product term in the RQL provision given by equation 86.

$$C_1\,(\mathrm{PD}_{VOIDS}) + C_2\,(\mathrm{PD}_{THICK}) + C_3\,(\mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK}) \;\geq\; 100 \qquad (86)$$

where: PD_{VOIDS} = air voids percent defective.
PD_{THICK} = thickness percent defective.
C_i terms = coefficients to be determined.

The threshold value of 100 on the righthand side of equation 86 that triggers the rejection provision is chosen arbitrarily and could be any convenient value. To determine the three coefficients C_1, C_2, and C_3, the three predetermined points in table 45 are substituted into equation 86 to obtain equations 87 through 89, thus providing three equations with three unknowns.

$$75\,C_1 + 10\,C_2 + 750\,C_3 = 100 \qquad (87)$$
$$10\,C_1 + 75\,C_2 + 750\,C_3 = 100 \qquad (88)$$
$$50\,C_1 + 50\,C_2 + 2500\,C_3 = 100 \qquad (89)$$

Solving these simultaneous equations, and substituting the values of the coefficients back into equation 86, produces equation 90, which is plotted as model #1 in figure 61.

$$1.273\ \mathrm{PD}_{VOIDS} + 1.273\ \mathrm{PD}_{THICK} - 0.0109\ (\mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK}) \;\geq\; 100 \qquad (90)$$

To demonstrate that the model can be made to bend the other way, if desired, and that greater weight can be put on one property, air voids for example, table 46 presents a slightly different set of assumptions that might have been used.

Table 46. Data Set Used to Develop Alternate RQL Provision Given by Equation 91

Quality Level		Rejectable ?
Air Voids	Thickness	
PD = 75	PD = 10	Yes
PD = 10	PD = 90	Yes
PD = 40	PD = 40	Yes

Solving the simultaneous equations generated by this data set produces equation 91, which has been plotted as model #2 in figure 61. By defining the rejectable level for thickness at a lower level of quality than the rejectable level for air voids, the coefficient of the thickness term in equation 91 has been reduced, thus giving air voids greater weight in this example.

$$1.076\ PD_{VOIDS} + 0.847\ PD_{THICK} + 0.0144\ (PD_{VOIDS} \times PD_{THICK}) \geq 100 \qquad (91)$$

Note that the coefficient of the cross product term in equation 90 is negative, producing a model that is concave–downward, while the positive cross product coefficient in equation 91 produces a concave–upward model. If there were no cross product term, i.e., if coefficient C_3 in equation 86 were zero, the model would plot as a straight line. Because an equation of this form can produce any of these three shapes, it can be very effective as a performance model when two quality characteristics are involved. The specific application will dictate which shape is appropriate.

Estimating Expected Pavement Lives

The method that is discussed in this and following sections is applied to the example of using air voids and thickness as acceptance measures for HMAC pavements. However, the concept that is presented is appropriate for both HMAC and PCC and for other acceptance measures, provided a method exists for estimating the pavement lives for various levels of the quality measure.

A highway agency can use whatever model or other method with which it is comfortable to arrive at the estimated lives for the as–constructed pavements. The methods used herein are only examples of possible approaches that can be used. If a performance model has been developed by the agency, then it may be possible to use this model to directly arrive at the expected pavement life for any combination of values for the variables included in the model.

Analytical Data

Returning to the example using air voids and thickness, to derive a mathematical performance model it will be necessary to have reasonably accurate values of expected life for the four conditions indicated in table 47. The values in this table were obtained by an agency using a simplified computer model that it has developed. The first value is obtained by assuming that the expected life of the pavement will equal the design life of 20 years if both air voids and thickness are at the AQL of 10 PD. Next, using the results obtained with the agency’s computer model, expected lives of approximately 10 years each are obtained for the cases in which either air voids or thickness is at the indicated poor level of quality while the other measure is at the AQL. Finally, a method must be found to estimate the joint effect of poor quality in both air voids and thickness to be able to complete the table. In the absence of actual data obtained under controlled field conditions, the agency decided that a survey of experienced pavement engineers would be necessary to estimate this missing piece of information.

Table 47. Preliminary Performance Matrix of Expected Life Values for HMAC Pavement Under NJDOT Conditions

Air Voids Quality	Thickness Quality	
	PD = 10	PD = 90
PD = 10	20 yrs	10 yrs
PD = 75	10 yrs	?

Survey Data

Figure 62 shows a completed survey questionnaire of the type that was sent by the agency to the chief engineer (or equivalent position) of all State transportation departments. A brief cover letter described the purpose of the survey and requested that it be forwarded to those individuals having extensive experience in the performance of HMAC pavements. Respondents were asked to estimate the expected life for seven different combinations of pavement quality under the assumption that acceptable quality in all three measures would result in the pavement providing the design life of 20 years.

Responses were received from 35 States, of which 4 indicated that this information was not available. Of the remaining responses, another five were excluded because some of the estimates of expected life were inconsistent with the assumption that, in a rational model, a large decrease in quality of any one parameter with the other parameters held constant would result in a corresponding decrease in expected life (provided it was not already zero). This left a total of 26 responses for the analysis, the averages of which appear on the survey questionnaire in figure 62.

The two matrices on the survey questionnaire provide two opportunities to examine how the effects of deficiencies in air voids and thickness should be combined based on the responses. Three different approaches were examined — additive, average, and product models — and the results are presented in table 48. It can be seen for case 1 in figure 62 that the effect of a change from good to poor in air–voids quality, with the other quality levels held constant, can be expressed as a decrease of 20 – 11.6 = 8.4 years, or as a ratio of expected life to design life of 11.6 / 20 = 0.58. Similarly, the effect due to thickness alone in case 1 is –5.0 years or a ratio of 0.75. Similar results are obtained for case 2.

If the effects were truly additive, the predicted life resulting from poor quality in both air voids and thickness for case 1 would be 20 – 8.4 – 5.0 = 6.6 years, as indicated in the sixth column in table 48. If the effects were averaged, the predicted life would be 20 – (8.4 + 5.0) / 2 = 13.3 years, which appears in the seventh column of table 48. By the product method, the predicted life would be the product of the individual ratios and the design life, or 0.58 × 0.75 × 20 = 8.7 years, which appears in the eighth column of table 48. For case 2, the average response for good quality of 16.1 is used in place of the design life.

HMA PAVEMENT PERFORMANCE SURVEY

Instructions: Please fill in the estimate of expected life (years) in the seven empty boxes using the assumptions listed at the bottom of the page.

CASE 1: GOOD Initial Smoothness		THICKNESS	
		GOOD	POOR
IN - PLACE AIR VOIDS	GOOD	20	15.0
	POOR	11.6	8.7

CASE 2: POOR Initial Smoothness		THICKNESS	
		GOOD	POOR
IN - PLACE AIR VOIDS	GOOD	16.1	11.9
	POOR	9.3	6.8

Assumptions:

- GOOD quality in all three categories (air voids, thickness, smoothness) corresponds to an expected service life (time after which resurfacing is required) of 20 years.
- For in-place air voids:
  GOOD = average value around 5 - 6 percent
  POOR = average value around 10 - 11 percent
- For thickness:
  GOOD = average value somewhat greater than design HMA layer thickness of 6 inches
  POOR = average thickness 1/2 to 3/4 inch less than the design value
- For smoothness (qualitative):
  GOOD = result that agency regards as clearly acceptable
  POOR = result that agency would accept only with substantial pay reduction

Figure 62. Average Results Obtained with Survey Questionnaire for HMA Pavement Performance

Table 48. Comparison of Three Methods for Combining Effects of Multiple Deficiencies

Case	Effect on Expected Life Due to Change from Good to Poor Quality [1]				Combined Predicted Life, yrs (Three Combining Methods)			Expected Life Based on Survey, yrs.
	Air Voids		Thickness					
	Years	Ratio	Years	Ratio	Add	Average	Product	
1	–8.4	0.580	–5.0	0.750	6.6	13.3	8.7	8.7
2	–6.8	0.578	–4.2	0.739	5.1	14.5	6.9	6.8

[1] Computed from survey results in figure 62

To judge which method is most appropriate, the last column of this table lists the average values estimated by the respondents of the survey. By comparing the values for predicted life with those in the last column, it is seen that the product method produces an almost exact agreement with the survey values, indicating that this method provides a good approximation of the manner in which experienced engineers believe the effects of multiple deficiencies manifest themselves. The average model greatly underestimates the expected loss of service life in this example, while the additive model, although overestimating the expected loss of service life, produces estimates that are reasonably close to the survey results.

These results suggest how a reasonable estimate for the missing value in the performance matrix in table 47 can be obtained. The values in table 47 indicate that poor quality in either air voids or thickness results in a ratio of expected life to design life of 10 / 20 = 0.50. Therefore, by the product method, the expected life when both air voids and thickness are at the indicated poor values is 0.50 × 0.50 × 20 = 5 years, completing the performance matrix as shown in table 49. Based on these four values for expected life, it is possible to develop a realistic performance model, and also to determine the appropriate equation form for the RQL provision discussed earlier.

Word of Caution

It is once again stressed that the above approach is simply one example of how these estimated pavement lives could be obtained. Any method with which the agency is comfortable can be used to develop the values for estimated pavement resulting from various levels of the quality measure. For example, if a performance model is available, and the highway agency has confidence in the predictive capability of the model, then it could be used to develop the estimated pavement expected lives. As noted, such performance models are under development but may be a number of years away from widespread use. These models tend to be quite complicated, and will not likely use typical quality measures such as PWL or PD as input variables.

Table 49. Final Matrix of Expected Life Values Used to Develop HMAC Pavement Performance Model

Air Voids Quality	Thickness Quality	
	PD = 10	PD = 90
PD = 10	20 yrs	10 yrs
PD = 75	10 yrs	5 yrs

Developing the Performance Model

Equation 92, patterned after the general form for the RQL provision in equation 86, is a practical model for a performance equation based on two quality characteristics. The expected pavement life in years is designated by *EXPLIF*, and all other terms are as previously defined.

$$EXPLIF = C_0 + C_1\,(\mathrm{PD}_{VOIDS}) + C_2\,(\mathrm{PD}_{THICK}) + C_3\,(\mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK}) \qquad (92)$$

The values for expected life in table 49 are used to develop four simultaneous equations that can be solved to provide the four equation coefficients. These are then substituted back into equation 92 to produce the performance model given by equation 93.

$$EXPLIF = 22.9 - 0.163\,\mathrm{PD}_{VOIDS} - 0.135\,\mathrm{PD}_{THICK} + 0.000961\,(\mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK}) \qquad (93)$$

It can be seen by inspection that this equation predicts that excellent quality ($\mathrm{PD}_{VOIDS} = \mathrm{PD}_{THICK} = 0$) will extend pavement life beyond its design life of 20 years to almost 23 years. It can also be readily calculated that the worst possible quality level in terms of percent defective ($\mathrm{PD}_{VOIDS} = \mathrm{PD}_{THICK} = 100$) produces an expected life of 2.7 years. Both of these are reasonable results based on the information that is available. As a further check of the derivation of equation 93, the four combinations of quality levels listed in table 49 can be entered into this equation to demonstrate that it returns the table 49 values for expected life.

For a clearer picture of the operation of this performance model, figure 63 has been plotted. Here it can be seen that, based on the assumptions listed in table 49, the appropriate shape for the family of curves is concave–downward. This indicates that, of the two possible models for an RQL provision shown in figure 61, the concave–downward model should be selected. If, for example, the agency decided that an expected life of 10 years, or less, was sufficiently detrimental that it warranted outright rejection, then the value of *EXPLIF* = 10 would be substituted into equation 93, and the results scaled accordingly, to produce the RQL provision given by equation 94. (Equation 94 is very similar to the RQL provision given by equation 90 and plotted as model #1 in figure 61. It is represented by the 10–year–lifeline in figure 63.)

$$1.264\,\mathrm{PD}_{VOIDS} + 1.047\,\mathrm{PD}_{THICK} - 0.00745\,(\mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK}) \geq 100 \qquad (94)$$

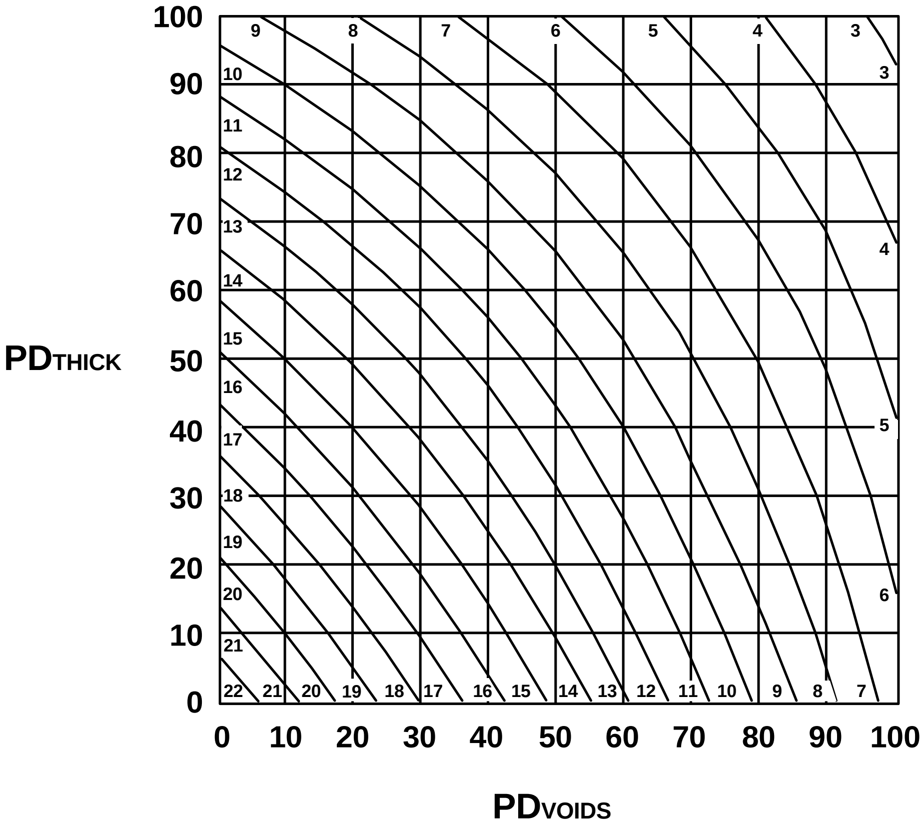


Figure 63. Graph of Expected Life, in Years, Generated by Equation 93

Summary and Conclusions

A method was presented illustrating how analytical and/or survey data can be used to develop a mathematical model to predict pavement performance as a function of acceptance test results. The method involved developing a simple matrix of expected life values that are used to construct a set of simultaneous equations that are solved to derive a simplified practical performance model.

This appendix also contains a summary of a nationwide survey conducted to determine the appropriate way to estimate the combined effect of multiple deficiencies. For this particular combination of quality characteristics (air voids and thickness), the analysis suggests that the combined effect is close to the sum of the individual effects, and appears to be best represented by the product of the ratios of the individual effects. A third method, based on the average of the individual effects, substantially underestimated the expected loss of service life.

APPENDIX K

Development of Composite Quality Measures

Background

Statistical construction specifications use characteristics measured at the jobsite to make appropriate acceptance decisions. The decision might be to pay a small incentive for unusually high quality, pay the contract price for an item that meets the specified quality level, accept the item at reduced payment when the quality is marginally deficient, or require removal and replacement if the quality is seriously defective. It may also be necessary to make an intermediate decision concerning whether or not additional tests should be performed to check marginal results, or to confirm that the quality is sufficiently poor that removal and replacement is warranted.

There are, therefore, several different types of decisions to be made and, when there are also different types of tests to be performed on a particular construction item, it can become a complex matter to design an acceptance procedure that is fair, effective, and free from inconsistencies. As an example of one type of inconsistency that can occur, consider the performance of a rejection provision for air voids and thickness of HMAC pavement as outlined in table 50.

Table 50. Example of an Inconsistent Rejection Provision

Case	Quality Level		Rejectable?
	Air Voids	Thickness	
1	PD = 10 (AQL)	PD = 75 (RQL)	Yes
2	PD = 75 (RQL)	PD = 10 (AQL)	Yes
3	PD = 74 (Almost RQL)	PD = 74 (Almost RQL)	No

The RQL for both air voids and thickness has been defined as PD ≥ 75. In other words, if either characteristic exhibits this level of quality, the lot may be declared rejectable. However, as seen in table 50, case 3 is clearly the worst case, yet it does not trigger the RQL provision whereas the other two cases do.

To correct this inconsistency, a composite RQL provision can be derived as a joint function of the two quality measures. Equation 95 and figure 64 illustrate an RQL provision of this type derived from basic performance considerations (see appendix J).

$$1.273\ \mathrm{PD}_{VOIDS} + 1.273\ \mathrm{PD}_{THICK} - 0.0109\ (\mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK}) \geq 100 \qquad (95)$$

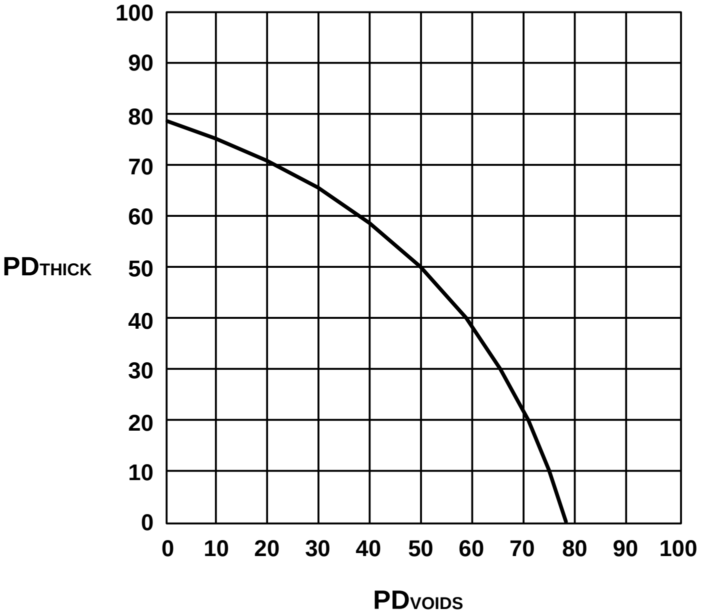


Figure 64. Graph of RQL Provision Given by Equation 95

All points on the curve in figure 64 are judged to be equally detrimental in terms of pavement performance, and any combination of PD_{VOIDS} and PD_{THICK} falling on or above the curve would be judged to be rejectable. Although equation 95 was developed specifically to make an appropriate rejection decision based on the joint quality measures of air voids and thickness, it suggests how a general family of curves might be developed that would be useful for the retest and payment adjustment decisions that must also be made.

To illustrate how this can be done, equation 96 presents a general performance model in which *EXPLIF* represents the expected life of the pavement in years, the *C* terms are coefficients to be determined, and PD_{VOIDS} and PD_{THICK} are the measures of air–voids and thickness quality. To determine the four unknown coefficients in equation 96, four known points spanning a wide range of quality are needed. These may be obtained from any valid source—agency experience, established performance models, or field experiments—and the values presented in table 51 were obtained from a combination of these sources (see appendix J).

$$EXPLIF = C_0 + C_1\,(\mathrm{PD}_{VOIDS}) + C_2\,(\mathrm{PD}_{THICK}) + C_3\,(\mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK}) \qquad (96)$$

Table 51. Performance Values for Expected Life Used to Develop Equation 101

	PD_{THICK} = 10	**PD_{THICK} = 90**
PD_{VOIDS} = 10	20 years	10 years
PD_{VOIDS} = 75	10 years	5 years

Substituting the values from table 51 into equation 96 produces four equations in four unknowns given by equations 97 through 100:

$$C_0 + 10\,C_1 + 10\,C_2 + 100\,C_3 = 20 \tag{97}$$
$$C_0 + 75\,C_1 + 10\,C_2 + 750\,C_3 = 10 \tag{98}$$
$$C_0 + 10\,C_1 + 90\,C_2 + 900\,C_3 = 10 \tag{99}$$
$$C_0 + 75\,C_1 + 90\,C_2 + 6750\,C_3 = 5 \tag{100}$$

The solution for this set of simultaneous equations provides the coefficients to be substituted into equation 96 to obtain the performance model given by equation 101:

$$EXPLIF = 22.9 - 0.163\ \mathrm{PD}_{VOIDS} - 0.135\ \mathrm{PD}_{THICK} + 0.000961\ (\mathrm{PD}_{VOIDS} \times \mathrm{PD}_{THICK}) \tag{101}$$

This equation predicts that excellent quality ($\mathrm{PD}_{VOIDS} = \mathrm{PD}_{THICK} = 0$) will extend pavement life beyond its design life of 20 years to almost 23 years. At the other extreme, the worst possible quality level in terms of percent defective ($\mathrm{PD}_{VOIDS} = \mathrm{PD}_{THICK} = 100$) produces an expected life of just less than three years. As a further check of the derivation of equation 101, the four combinations of quality levels listed in table 51 can be entered into this equation to demonstrate that it returns the table 51 values for expected life. The complete family of performance curves is plotted in figure 65.

Because equation 101 provides a direct estimate of expected life as a function of the individual quality measures, it can form the basis for a performance–related payment schedule, as demonstrated in chapter 6. Alternatively, equation 101 can be used to develop a single, composite quality measure that will simplify the overall acceptance process and the various decisions to be made.

Since the quality measure used for this example is percent defective (PD), which ranges from a minimum of zero to a maximum of 100, it will be convenient to develop a composite quality measure (PD*) that spans that same range. As derived, the value of *EXPLIF* in equation 101 ranges from 2.71 to 22.9. To map equation 101 onto the desired response surface, the following three operations must be performed on the terms on the right–hand side of the equation:

- Subtract 2.71.
- Multiply by 100 ÷ (22.9 – 2.71) = 4.953.
- Subtract the resulting equation from 100.

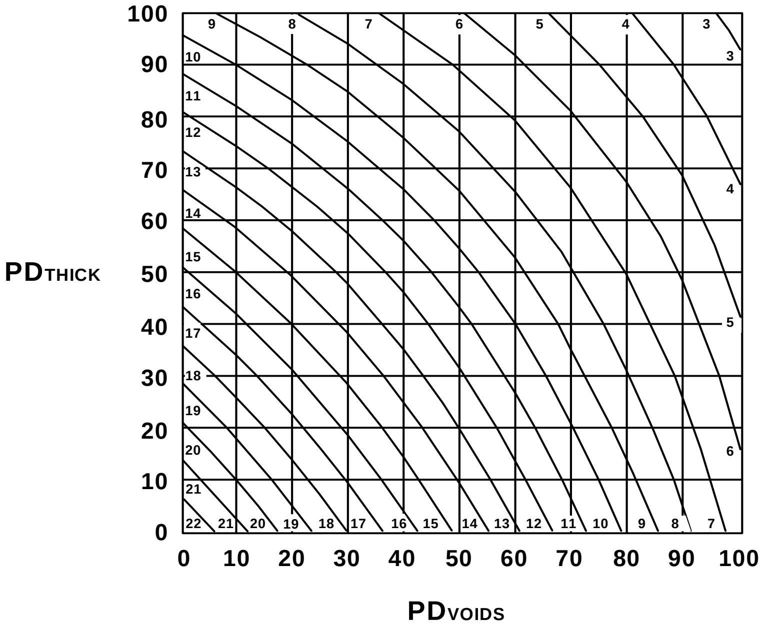


Figure 65. Graph of Expected Life in Years Generated by Equation 101

This produces equation 102, which can be demonstrated to produce values of PD* ranging from zero to 100 as the individual quality measures cover the entire range from $PD_{VOIDS} = 0$, $PD_{THICK} = 0$ to $PD_{VOIDS} = 100$, $PD_{THICK} = 100$. To provide a better understanding of how this equation performs, a family of curves is plotted in figure 66.

$$PD^* = 0.807\ PD_{VOIDS} + 0.669\ PD_{THICK} - 0.00476\ (PD_{VOIDS} \times PD_{THICK}) \qquad (102)$$

Practicality of Composite Quality Measures

To demonstrate the practicality of the composite quality measure, a complete acceptance procedure must be specified. This includes the AQL, the RQL, any retest provision, and the payment schedule.

For illustration purposes, suppose the agency has decided that the AQL, the quality level at which the pavement is to be accepted with no price adjustment, is defined as $PD_{VOIDS} = PD_{THICK} = 10$. Suppose, also, that if the combined effects of poor quality in air voids and thickness are such that the performance relationship of equation 101 (shown graphically in figure 65) indicates an expected life of 10 years, or less, the pavement is considered to be rejectable and the agency reserves the option to require removal and replacement. Also, since agencies might want to specify some intermediate level of quality at which the agency has the option to require additional tests before accepting what might turn out to be seriously defective work, it will be assumed for this example that the retest option is in effect whenever the performance model of equation 101 indicates an expected life of 15 years or less. And finally, based on a life–cycle cost analysis of the cost of premature pavement failure,(32) the projected net present cost to the highway agency of a 20–year design pavement failing after only 10 years is estimated to be approximately $50,000 / lane–kilometer.

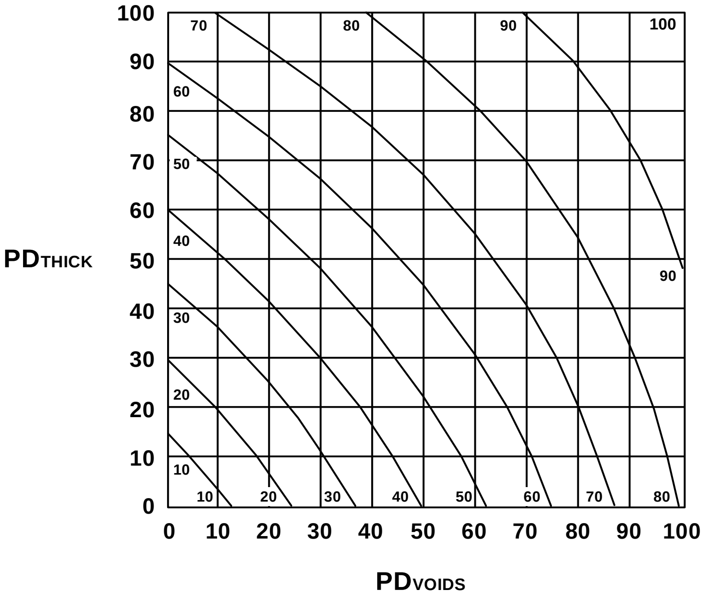


Figure 66. Graph of Composite Quality Measure (PD*) Given by Equation 102

With this information, the complete acceptance procedure can be specified. To determine the comparable value of PD* associated with the AQL, the values of $\mathrm{PD}_{VOIDS} = \mathrm{PD}_{THICK} = 10$ are substituted into equation 102 to obtain PD* = 14 as the AQL. Therefore, the payment equation must produce a payment adjustment of zero at PD* = 14. To determine the value of PD* associated with the RQL, any combination of values on the 10–year–lifeline in figure 65 can be substituted into equation 102. For example, entering $\mathrm{PD}_{VOIDS} = 60$ and $\mathrm{PD}_{THICK} = 40$ into equation 102 produces $\mathrm{PD}^* \geq 64$ as the RQL. Similarly, any combination of values on the 15–year–lifeline, such as $\mathrm{PD}_{VOIDS} = 20$ and $\mathrm{PD}_{THICK} = 40$, produces $\mathrm{PD}^* \geq 39$ as the retest provision. Finally, assuming a simple linear payment equation will be sufficient, the payment schedule given by equation 103 is derived.

$$PAYADJ = 14{,}000 - 1000\ \mathrm{PD}^* \qquad (103)$$

Where: $PAYADJ$ = lot payment adjustment ($ / lane–kilometer), and
PD* = composite quality measure.

It can be seen that when PD* is at the AQL value of 14, equation 103 produces a payment adjustment of zero, as desired. Similarly, when PD* equals the RQL value of 64, the desired payment reduction of –$50,000 / lane–kilometer is obtained. For truly excellent quality, where PD_{VOIDS} and PD_{THICK} are both zero, PD* also equals zero and the payment equation awards a maximum bonus of $14,000 / lane–kilometer. At the other extreme, when PD_{VOIDS} and PD_{THICK} both equal 100, PD* also equals 100 and the payment equation assigns the maximum payment reduction of –$86,000 / lane–kilometer. In between, all payment adjustments are related to performance in that all quality levels that fall on any particular pavement–lifeline in figure 65 will receive the same level of payment. Note also that both the retest provision and the RQL provision, because they are stated in terms of the composite quality measure, avoid the inconsistency illustrated in table 50.

Optional Compound Payment Equation

The maximum incentive of $14,000 / lane–kilometer paid by equation 103 is the result of choosing to make the payment equation linear. If the agency were not comfortable with this large an incentive, perhaps because there were other factors that might prevent the pavement from fully benefiting from superior quality in the measured quality characteristics, then it may be appropriate to use a compound payment equation that is less steeply inclined in the region of the AQL. For example, suppose the agency determined that the maximum incentive to be paid under any circumstances was $5000 / lane–kilometer, but that the payment reduction of –$50,000 / lane–kilometer at PD* = 64 was still considered to be justified. Since it is necessary to produce a payment adjustment of zero at the AQL of PD* = 14, two separate payment equations will be required, resulting in a compound payment equation. Although the breakpoint of the compound equation can be at any desired point, it will be convenient to put it at the retest level of PD* = 39. The resulting payment schedule is given by equations 104 and 105, and is illustrated in figure 67.

$$\text{PD*} < 39: \qquad PAYADJ = 5{,}000 - 357\ \text{PD*} \tag{104}$$

$$\text{PD*} \geq 39: \qquad PAYADJ = 55{,}154 - 1643\ \text{PD*} \tag{105}$$

Extension to Three Variables

Although it is possible to apply this same process when there are three individual quality characteristics, more data are required and it will be necessary to solve eight simultaneous equations to account for all the cross-product terms. Therefore, when there are three or more variables, it is recommended that the exponential model, which is described in chapter 6 and appendix L, be used.

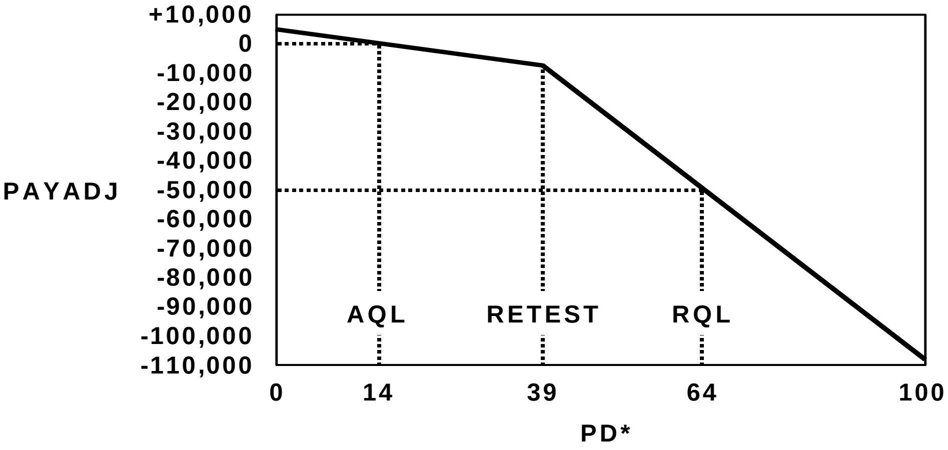


Figure 67. Compound Payment Schedule Given by Equations 104 and 105

Suitability of Composite Measures

Not all individual quality measures are equally suitable for incorporation into a composite measure. Measures that are best suited are those that jointly affect performance in such a way that, within practical limits, higher quality in one tends to offset deficiencies in the others. Another requirement is that they be convenient to measure in association with each acceptance lot. The example involving air voids and thickness of HMAC pavement is obviously well suited, as would be strength and thickness of PCC pavement.

Measures that are less suitable are those that are dominant or controlling, or which are less convenient to measure on individual lots. For example, an agency might choose to use a stand–alone acceptance procedure for pavement smoothness if it were felt that riding quality should be evaluated on its own merits without regard to the other quality measures, or if it were considered more convenient to evaluate riding quality on large sections of the project at a single time rather than on a lot–by–lot basis.

Summary and Conclusions

An alternate method to develop performance–related payment equations for multiple quality characteristics is to first develop a composite quality measure that is a function of the individual quality measures. An example was presented in which performance relationships for air voids and thickness of HMAC pavement were used to develop a single quality indicator, PD*, as a function of the individual quality measures, PD_{VOIDS} and PD_{THICK}.

The methodology offers both mathematical simplicity and a sound empirical basis in that it can be developed from performance data obtained by a variety of means. It is generally applicable in any situation involving multiple acceptance criteria for which performance results are known or can be modeled or approximated. It is well suited for use with PD or PWL, but is not necessarily limited to those quality measures. It avoids certain inconsistencies in practice that may occur with other methods for dealing with multiple quality characteristics, and it leads to rational payment schedules in that it assures that all combinations of individual quality measures that predict the same level of expected life will receive the same amount of payment adjustment.

The use of a composite quality measure provides a practical and effective means to make three different types of acceptance decision—retest, reject, and payment adjustment—all keyed to a common performance model. But in spite of these many advantages, perhaps its strongest selling point is the ease with which it can be understood and applied, as demonstrated by the following procedural steps for the HMAC example:

- Obtain the air voids and thickness test results from the acceptance lot.
- Compute PD_{VOIDS} and PD_{THICK} in the customary manner.
- Compute $PD^* = 0.807\ PD_{VOIDS} + 0.669\ PD_{THICK} - 0.00476\ (PD_{VOIDS} \times PD_{THICK})$.
- Check $PD^* \geq 39$ for the retest option and make the appropriate decision.
- Check $PD^* \geq 64$ for the RQL option and make the appropriate decision.
- Compute $PAYADJ = 14{,}000 - 1000\ PD^*$ (or other suitable payment equation).

APPENDIX L

Model for Determining Payment When Using Multiple Quality Characteristics

Background

Statistically based construction specifications and acceptance procedures have been widely used by the highway profession for many years. These procedures typically specify an end result that can be measured in statistical terms and award payment in proportion to the extent to which the end result has been achieved. Most prescribe a variable level of payment reduction for varying amounts of deficient quality, and many now also include modest bonuses for quality that substantially exceeds the level that has been specified.

For such procedures to be considered equitable and defensible, they should not be punitive but should award payment that is at least approximately commensurate with the value received. To satisfy this requirement, it is necessary to base these procedures on quantitative (mathematical) models relating as–built construction quality to expected service life and value.

Consequently, current efforts tend to be directed at refining these procedures to become true PRS, specifications based on quality characteristics measured at the jobsite that can be related to the performance of the construction item in a specific, quantitative manner. For example, an obvious performance measure for highway pavement is service life. If, based on as–built quality levels, it is found that the pavement does not measure up to design standards, the relationship between construction characteristics and load–carrying capacity can be used to estimate the amount by which its service life will be shortened. This, in turn, can be combined with a LCC analysis to compute the expected loss in net present value, thus justifying an appropriate amount of payment reduction.

Obviously, a prerequisite for such a process is to first develop the mathematical models necessary to predict from quality characteristics measured at the jobsite what the expected performance of the construction item will be. In the case of pavement, the design procedure itself provides some of the necessary relationships. If the pavement is constructed with insufficient thickness, for example, it is a simple matter to work backwards through the design procedure to determine the reduced number of loads it will be capable of sustaining and, from that, its expected life can be predicted. For other characteristics, however, no such convenient relationships may exist and, in these cases, it is necessary to develop a method to obtain the required relationships from existing knowledge or data.

Although it is possible to obtain a mathematical model by performing a least–squares fit with a standard regression program, using either a linear equation or various curvilinear forms, it is often possible to obtain a better model by first using engineering knowledge to determine the most appropriate mathematical form. In many cases, using known boundary conditions to reason out the general shape of the mathematical relationship will provide a much improved model that is more likely to be accurate throughout its range, not just in the region in which most of the data are concentrated.

This appendix presents a straightforward and practical procedure by which any agency can make use of empirical performance data to develop quantitative models for expected life and value for multiple quality characteristics, thus forming the basis for rational and defensible payment adjustment schedules.

Need for Simplified Modeling Method

Although efforts are under way to create extremely sophisticated computerized procedures to develop performance relationships and appropriate payment schedules, the successful completion and validation of these procedures is still some time away. Even when completed, the data requirements and level of complexity of these procedures may deter their widespread use by practitioners seeking more practical methods that are easier to understand and to apply. Therefore, there is need for an alternative approach for those agencies that choose to develop their own procedures in their own way tailored to their own specific circumstances. Perhaps more importantly, this alternative method needs to be sufficiently straightforward and scientifically sound so that agency engineers can not only understand it and use it with confidence, but also can modify it when necessary and be able to present it convincingly to the contractors whose work it will govern. Without this degree of familiarity and understanding, it may not be easy to implement the new procedures in the face of the typical opposition from the construction industry, nor to explain or defend the results should they be challenged.

Basic Model Form

Suppose it were desired to develop a model for PSI based on the output of some measuring device that is either pushed or driven down the road. PSI is defined on a scale of 0.0 – 5.0 in which 5.0 represents perfect smoothness and 0.0 indicates a pavement that is virtually impassible. The 0.0 – 5.0 scale was initially developed for use by a team of raters to evaluate pavements of varying degrees of roughness but, to be useful as a practical evaluation tool, it is necessary to relate it to the output of some standardized mechanical measuring device.

To determine an appropriate mathematical form for this relationship, first consider the situation illustrated by point “A” in step 1 of figure 68. If a highly sensitive measuring device produces no output whatsoever (i.e., zero on the horizontal–axis), then by definition that corresponds to a PSI value of 5.0 on the vertical–axis (i.e., perfect smoothness). Next, consider point “B” at which the measuring device produces a very small reading, but the level of roughness is so slight that it is undetectable by a panel of raters. In this case, the pavement would still be rated at 5.0, even though there is a very small output from the measuring device. However, as the test pavements become rougher, both the measuring device and the panel will begin to record readings below 5.0. What this indicates is that the upper portion of the model very likely starts out horizontally from the vertical–axis at and then begins to bend downward, as indicated by the curved line through points “A” and “B” in figure 68.

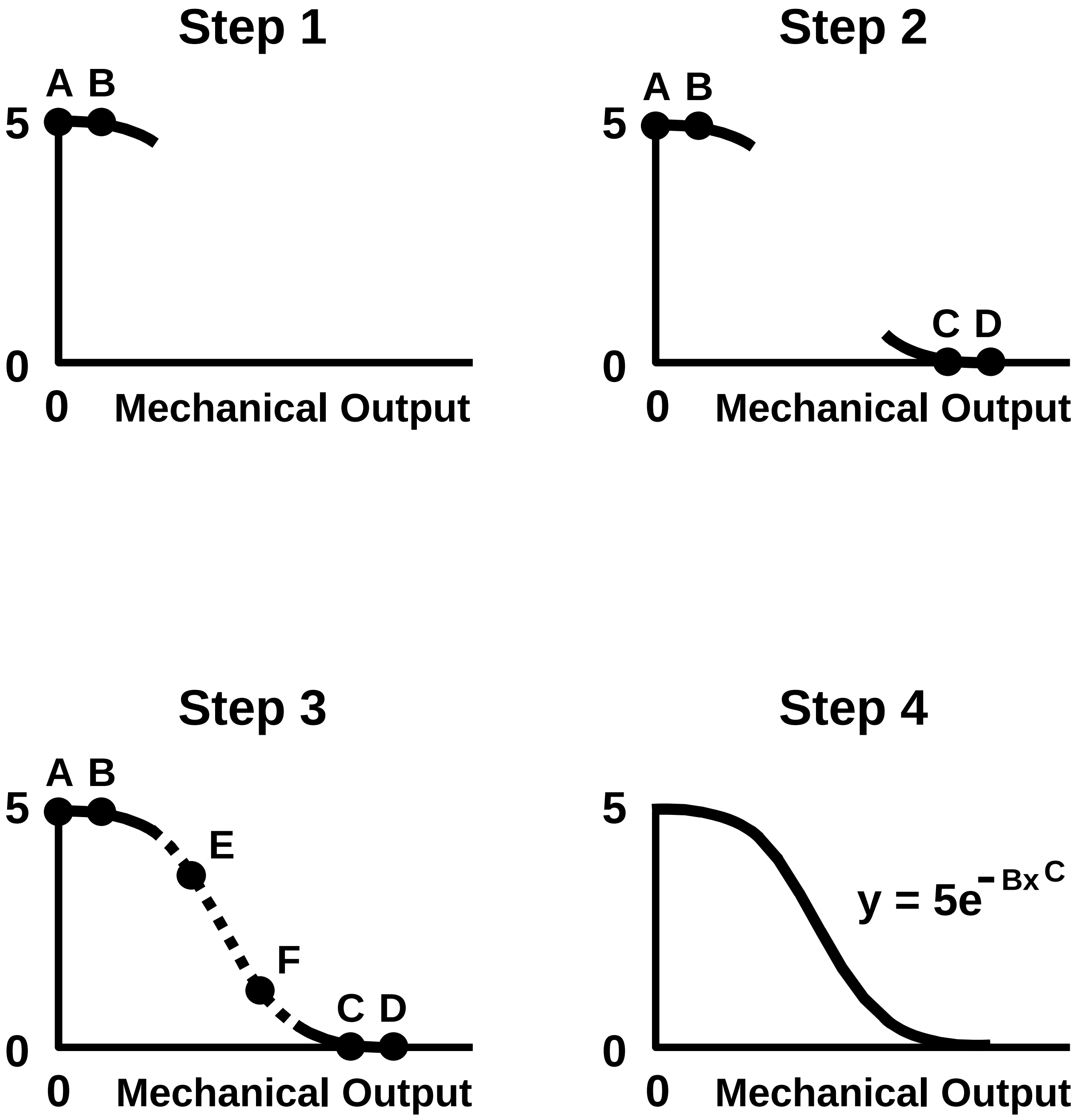


Figure 68. Conceptual Steps to Develop a Performance Model for Pavement Smoothness

Next, consider the lower end of the curve shown in step 2 of figure 68. It would be possible to construct a hypothetical pavement that is so rough that the rating team would consider it impassible and rate it a zero, indicated by point “C.” However, it would also be possible to construct a still rougher pavement as measured by the testing device, indicated by point “D,” and this pavement would also be rated a zero because that is the minimum possible rating. What this indicates is that the curve must also be horizontal at its lower extremity, and thus comes in asymptotically to the horizontal–axis.

These two boundary conditions make it possible to narrow the range of choices for the appropriate mathematical model to an equation that produces a sigmoidal (“S”) shape as shown in step 3 of figure 68. Then, if there are sufficient data in the middle region, or if one or two additional points can be identified, such as points “E” and “F” in step 3, the model can be reasonably well specified.

A convenient mathematical form to model this curve is the exponential expression given by equation 106 and shown in step 4 of figure 68. For a perfectly smooth pavement, $y = 5$ at $x = 0$, so the first coefficient is predetermined to be $A = 5$. The two “known” points “E” and “F” provide sufficient information to write two simultaneous equations that can be solved to obtain the remaining two coefficients in equation 106.

$$y = Ae^{-Bx^{C}} \tag{106}$$

There are two reasons why equation 106 does a good job of modeling performance relationships. The first is that it properly takes boundary conditions into account and will not return unrealistic values in the extreme regions for which data may not have been available. The second is that, since there are three unknown coefficients (A, B, and C) to be determined, this permits the model to be made to pass through three convenient known points, such as points “A,” “E,” and “F” in figure 68. Therefore, provided the level of performance can be estimated at three points, the entire model can be specified reasonably accurately. For a single–parameter model such as equation 106, three logical determining points would be the maximum (point “A” in figure 68), the AQL, and the RQL.

Selection of Statistical Quality Measure

To develop a general model, it will be desirable to replace the independent variable, x, in equation 106 with a suitable statistical quality measure. Of the various measures that might be chosen, probably the most frequently used for highway construction are PWL, or its complement, the PD. These two measures, which are functionally equivalent, are intuitively appealing because they account for both mean level and variability in a single statistic.

It can be observed that the model form given by equation 106 is especially well suited when zero is the most favorable level of x, as is the case when PD is substituted for x as the quality measure. When $x = 0$ is not the most favorable level, as is the case with PWL, a somewhat more complex equation form will be necessary. However, if it is desired to develop the model in terms of PWL, it will be found that it is better to first develop it in terms of PD, and then substitute (100 – PWL) for PD at the last step. Because the measure PD is the more natural fit for this equation form, it is used in the developments that follow.

Multiple–Parameter Model Form

Using equation 106 as a guide, the logical form for a multiple–parameter model is given by equation 107:

$$y = Ae^{-(B_1 x_1^{C_1} + B_2 x_2^{C_2} + \cdots + B_k x_k^{C_k})} \tag{107}$$

It is shown later, however, that a very serviceable model can be obtained without the need for the added complexity of the individual “C” exponents in equation 107, so they are omitted in the equation to be developed. Since PD is the more natural fit for the single–parameter model given by equation 106, it is selected as the independent variable for use with the multiple–parameter model. Finally, since it is desired to obtain an expression for expected service life as a function of as–built quality, the variable EXPLIF is substituted for *y* as the dependent variable, producing equation 108, in which *e* is the base of natural logarithms. (The final proof of this model is an extensive series of tests to demonstrate that it reliably produces results that are consistent with field experience.)

$$EXPLIF = Ae^{-(B_1 \mathrm{PD}_1 + B_2 \mathrm{PD}_2 + \cdots + B_k \mathrm{PD}_k)} \tag{108}$$

To solve for the unknown coefficients in equation 108, it is first necessary to take logarithms of both sides, producing equation 109:

$$\ln(EXPLIF) = \ln(A) - B_1 \mathrm{PD}_1 - B_2 \mathrm{PD}_2 - \cdots - B_k \mathrm{PD}_k \tag{109}$$

At this point, it is convenient to observe that the term ln(A) in equation 109 can be regarded as B_0, the constant term of the multiple–parameter expression. After a set of simultaneous equations has been solved to obtain the “B” coefficients, B_0 can be used to determine coefficient “A” using equation 110.

$$A = e^{B_0} \tag{110}$$

To summarize up to this point, equations 106 through 110 describe the mathematical basis for a practical multiple–parameter model. The next step is to outline the data requirements for this model.

Data Matrix

If the method is to be valid, it must be based on realistic data, and if it is to be practical, the required data must be readily obtainable. Table 52 is a generic data matrix that must be completed to apply this method.

Table 52. Data to Develop Multiple–Characteristic Performance Model

$\mathbf{PD_1}$	$\mathbf{PD_2}$	$\mathbf{PD_3}$	→	$\mathbf{PD}_k$	***EXPLIF***
AQL(1)	AQL(2)	AQL(3)	→	AQL(*k*)	*DESLIF*
POOR(1)	AQL(2)	AQL(3)	→	AQL(*k*)	*LIFE*(1)
AQL(1)	POOR(2)	AQL(3)	→	AQL(*k*)	*LIFE*(2)
AQL(1)	AQL(2)	POOR(3)	→	AQL(*k*)	*LIFE*(3)
↓	↓	↓	→	↓	↓
AQL(1)	AQL(2)	AQL(3)	→	POOR(*k*)	*LIFE*(*n*)

PD_i = percent defective for each of the *k* quality characteristics.
EXPLIF = expected life in years.
DESLIF = design life in years.
AQL(*i*) = acceptable quality level, in PD, for each of the *k* quality characteristics.
POOR(*i*) = poor quality level, in PD, for each of the *k* quality characteristics.
LIFE(*m*) = expected life in years for *n* selected combinations of PD levels.

It can be seen from the first row in table 52 that, when all quality characteristics are at their respective AQL values, the expected life is equal to the design life. For the remainder of the rows in this table, each characteristic in turn is set at some specified poor level of quality (which might appropriately be the RQL) while all the others are held constant at the AQL. It is believed that this provides the most convenient arrangement of performance data that an agency might be expected to have, or could obtain relatively easily. The values to be entered in the table might be developed as the collective opinion of experienced pavement engineers, or they might be obtained more formally through a multiple regression analysis of actual field data. In some cases, the agency's current pavement design method may be able to provide some of this information.

The next section demonstrates how easy it is to convert the data in this matrix to a performance model of the general form given by equation 108.

Illustrative Example

For demonstration purposes, consider a specification for HMAC pavement for which the agency wishes to control three quality characteristics: in–place air voids, thickness, and smoothness. A typical value for the AQL is PD = 10, while RQL values tend to vary more widely, depending on what level of quality an agency believes justifies potential removal and replacement. For this example, the values listed in table 53 have been selected.

Table 53. AQL and RQL Values Selected for the Example

Quality Characteristic	**PD (AQL)**	**PD (RQL)**
Air Voids	10	65
Thickness	10	75
Smoothness	10	85

The values in table 53, plus the design life for a typical overlay of 10 years, are entered into the general matrix of table 52, producing table 54:

Table 54. Preliminary Performance Matrix for the Example

PD_{VOIDS}	**PD_{THICK}**	**PD_{SMOOTH}**	***EXPLIF*, in years**
10 (AQL)	10 (AQL)	10 (AQL)	10 (*DESLIF*)
65 (RQL)	10 (AQL)	10 (AQL)	*LIFE* (poor voids)
10 (AQL)	75 (RQL)	10 (AQL)	*LIFE* (poor thickness)
10 (AQL)	10 (AQL)	85 (RQL)	*LIFE* (poor smoothness)

By now, the ease of applying this method should be apparent. The only additional pieces of information required are realistic estimates of expected service life for the three conditions specified in the last three rows of table 54, using any of the methods suggested in the previous section. For purposes of this example, assume that the agency has selected the respective individual RQL values in the belief that they will produce a loss of service life of about 50 percent, producing an expected life of 5 years for each of these rows. The final performance matrix is presented in table 55.

Table 55. Final Performance Matrix for the Example

PD_{VOIDS}	**PD_{THICK}**	**PD_{SMOOTH}**	***EXPLIF*, in years**
10	10	10	10
65	10	10	5
10	75	10	5
10	10	85	5

All that remains now is to substitute the information from the final performance matrix in table 55 into equation 109 to produce the necessary set of simultaneous equations. These are presented as equations 111 through 114:

$$2.302585 = B_0 - 10B_1 - 10B_2 - 10B_3 \tag{111}$$

$$1.609440 = B_0 - 65B_1 - 10B_2 - 10B_3 \tag{112}$$

$$1.609440 = B_0 - 10B_1 - 75B_2 - 10B_3 \tag{113}$$

$$1.609440 = B_0 - 10B_1 - 10B_2 - 85B_3 \tag{114}$$

These equations could be solved by hand, but there are any number of computer packages that will do this, producing the following results:

$$B_0 = 2.627669$$

$$B_1 = 0.012603$$

$$B_2 = 0.010664$$

$$B_3 = 0.009242$$

Then, using equation 110,

$$A = e^{B_0} = 13.84147$$

All the necessary constant terms have now been determined and the complete performance model can be written as equation 115.

$$EXPLIF = 13.8e^{-(0.0126\mathrm{PD}_{VOIDS} + 0.0107\mathrm{PD}_{THICK} + 0.00924\mathrm{PD}_{SMOOTH})} \tag{115}$$

Checking The Model

The first test of equation 115 is to check that it returns precisely the values from table 55 that were used to derive it. These results are presented in table 56.

Table 56. Check of the Derivation of Equation 115

Values Entered into Equation 115			***EXPLIF* Returned, in years**
PD$_{VOIDS}$	**PD**$_{THICK}$	**PD**$_{SMOOTH}$	
10	10	10	10.0
65	10	10	5.0
10	75	10	5.0
10	10	85	5.0

It is seen from table 56 that the model survives the first test because it returns exactly the appropriate values. A second test is to check the extremes, an area in which many models break down. The extremes in this case occur when all three PD values are either 0 or 100. These results are presented in table 57.

Table 57. Test of Equation 115 at Extreme Values

Values Entered into Equation 115			***EXPLIF* Returned, in years**
PD$_{VOIDS}$	**PD**$_{THICK}$	**PD**$_{SMOOTH}$	
0	0	0	13.8
100	100	100	0.5

Here again, the values returned by the performance model in equation 115 appear to be appropriate. It is not unreasonable to expect that a few exceptionally well constructed overlays may last 14 years, and in some cases, even longer. Therefore, the prediction by this model that the highest possible quality level would lead to an expected life of nearly 14 years seems reasonable. At the other extreme, the failure of a pavement during the first year is certainly rare, but it has occurred, so the prediction that the worst possible quality level (100 percent defective in all characteristics) could produce such an early failure may well be realistic. At this stage, the model is judged to be believable, but several additional tests are required.

The predicted life for a wide variety of combinations of individual quality levels of the three characteristics is presented in table 58. The first group of tests provides a sense of how expected life decreases as the three quality measures decline together. The second set of tests in this table illustrates how extra quality in some characteristics can offset deficient quality in other characteristics, all producing the design life of 10 years. This is an inherent feature in most design methods, and is believed to be an appropriate feature in any model of multiple characteristics. The only concern would be if extremely poor quality (100 percent defective) in one characteristic could be masked by superior quality in other characteristics, and the third group of tests indicates this is not the case.

Table 58. Additional Tests of Equation 115

Values Entered into Equation 115			***EXPLIF* Returned, in years**
PD_{VOIDS}	**PD_{THICK}**	**PD_{SMOOTH}**	
0	0	0	13.8
10	10	10	10.0
25	25	25	6.1
50	50	50	2.7
75	75	75	1.2
100	100	100	0.8
17	10	0	10.0
0	21	10	10.0
0	10	23	10.0
25	0	0	10.1
0	30	0	10.0
0	0	35	10.0
0	0	100	5.5
0	100	0	4.7
100	0	0	3.9
0	0	50	8.7
0	50	0	8.1
50	0	0	7.3

The final group of tests in table 58 is included to investigate the extent to which moderately poor quality (50 percent defective) could be offset by excellent quality in the other two characteristics. These three cases are examined individually, beginning with the smoothness case in the third row from the bottom in table 58. Using an upper specification limit of an IRI of 1.18 m/km and a typical standard deviation based on one agency's data of about $\sigma = 0.24$ m/km, the largest IRI value in a section of pavement having PD = 50 would be about three standard deviations above the limit, or about 1.90 m/km. According to recent literature on pavement profiling, (26) even new pavements may range up to about IRI = 3.16 m/km, so a newly constructed pavement having PD = 50 should have a considerable amount of service life remaining. Therefore, the expected life of 8.7 years predicted by equation 115 in table 58 appears to be reasonable.

In the next to last row of table 58, it is seen that a thickness quality level of PD = 50, combined with PD = 0 in the other two characteristics, produces an expected life of 8.1 years. To check this case, the AASHTO Design Procedure for Flexible Pavement is used. (31) Using a typical 100–mm overlay and nominal values for the various other variables (layer coefficient, resilient modulus, terminal serviceability, thickness standard deviation, etc.), it was calculated that a decrease in thickness quality from the AQL of PD = 10 to a moderately defective level of PD = 50 would result in a loss of load–carrying capacity of about 40 percent. Based on the design life of 10 years, the resultant life expectancy would then be about 6 years, somewhat lower than the value of 8.1 years in table 58. However, this calculation with the AASHTO design procedure assumes that other variables that are not included in the design procedure are at nominal satisfactory levels, whereas the example in table 58 has both air voids and smoothness at the best possible quality level of PD = 0. Therefore, it is logical to assume that this might raise the overall performance somewhat, and the predicted life of 8.1 years may be reasonable after all.

To perform a rough check on the last row of table 58, published information on the effects of HMAC compaction is used. It has been reported that the expected life of HMAC pavement decreases by approximately 10 percent for each 1.0 percent increase in the level of air voids above 7.0 percent. (33) Based on typical data from one agency, a decrease in quality from the acceptable level of PD = 10 to a defective level of PD = 50 requires a shift in average level of about 2.0 to 3.0 percent, so 2.5 percent is used here. Using the relationship cited above, this would correspond to a loss of life of 25 percent, or an expected life of about 7.5 years. This might appear to be in close agreement with the value of 7.3 years in table 58, but, as in the previous calculation, this does not account for the potentially beneficial effect of excellent quality in the other two variables, so the true value may be higher. However, since the information used to perform this check is only approximate, this may still be a reasonably close check for practical purposes.

Table 58 also provides the opportunity to observe the effects of changes in the individual quality characteristics. Going back to the values in the performance matrix in table 55, it can be seen that, based on the data used in this example, air voids has the greatest influence because it requires the smallest amount of percent defective (PD = 65) to reduce the expected life to 5 years, while smoothness has the least effect because it requires the greatest amount of percent defective (PD = 85) to produce the same effect. The second, third, and fourth groups of data in table 58 all demonstrate the consistency of the relative importance of these three variables. If a different combination of relative importance were desired, then different values would be used in the final performance matrix in table 55.

In summary, all of the checks of the model given by equation 115 have shown it to be both reasonable and consistent. More importantly, this method will reliably produce models that will accurately and consistently reflect the information entered into the performance matrices from which they are derived.

Converting Expected Life Into Value

The next step of this process is to determine the economic impact of the estimate of expected life obtained with the performance model. A practical repair strategy for HMAC pavement is particularly well suited for LCC analysis that can be used to properly discount future expenses. For at least some agencies, it is not normal practice to perform special maintenance actions just to restore the design life of a HMAC pavement that was constructed with some sort of quality deficiency. (An obvious exception would be a safety issue.) Instead, the condition of the pavement is monitored and, when premature failure begins to occur, the pavement is scheduled for an overlay. The availability of a performance model to predict when this premature failure will occur is obviously a critical component of an acceptance scheme designed to award payment in proportion to expected performance.

On the assumption that an estimate of expected life is available, and that it is justifiable to assign a payment reduction equivalent to the loss in net present value resulting from premature failure as the result of insufficient quality of items under the contractor's control, equation 116 is derived in appendix I:

$$PAYADJ = \frac{C(R^D - R^E)}{(1 - R^O)} \tag{116}$$

where: $PAYADJ$ = appropriate payment adjustment for new pavement or overlay (same units as C).

C = present total cost of resurfacing. (typical value = \$23.92/m^2).

D = design life of pavement or initial overlay (typically 20 years for new pavement, 10 years for overlay).

E = expected life of pavement or overlay (variable).

O = expected life of successive overlays (typically 10 years).

R = $(1 + INF) / (1 + INT)$.

INF = long–term annual inflation rate in decimal form (typically 0.04).

INT = long–term annual interest rate in decimal form (typically 0.08).

Table 59 has been constructed to show that equation 116 justifies relatively large payment adjustments that reflect real costs (or benefits) to the agency when the actual quality differs substantially from the design quality. Many possible combinations of quality levels are included, arranged in descending order from best to worst. Note that, although appropriate payment levels have been computed for all cases, many agencies would choose to have an RQL provision supersede the payment schedule for extremely low values of expected life, providing the option to require removal and replacement at the time of construction.

It is believed that few, if any, agencies use payment adjustments as large as those computed in table 59, possibly because they have lacked a firm basis to justify values this large. However, another explanation could be that it often is possible to get the desired response from the construction industry without using the maximum amount of payment adjustment that would be economically justifiable. In other words, the level of payment adjustment only needs to be large enough to provide a strong incentive to the contractor to produce good quality initially. The

Table 59. Range of Values Computed with Equations 115 and 116

Individual Quality Levels			Expected Life, in years	Payment Adjustment [1], in $/Lane Kilometer
PD$_{VOIDS}$	**PD**$_{THICK}$	**PD**$_{SMOOTH}$		
0	0	0	13.8	+25,484
5	0	5	12.4	+16,517
5	5	5	11.8	+12,527
10	10	10	10.0	0
0	0	45	9.1	–6,590
0	45	0	8.6	–10,349
45	0	0	7.9	–15,732
25	15	30	6.5	–26,935
40	15	30	5.4	–36,162
40	30	30	4.6	–43,117
40	30	55	3.6	–52,112
65	30	55	2.7	–60,502
65	75	55	1.6	–71,152
90	90	90	0.7	–80,202

[1] Computed using equation 116 and associated constants.

significance of this is that, if payment schedules substantially less severe than those that would be justifiable are typically used, then the performance models upon which the payment schedules are based do not have to be known with great precision. Therefore, models developed by the method outlined in this appendix are believed to be more than adequate for their intended use.

Developing the Payment Schedule

The final step of this process is to convert the information in table 59 into a workable payment equation. The easiest way to do this is to first plot the appropriate payment adjustment from table 59 versus expected life, as shown in figure 69. This relationship plots so nearly as a straight line that it could be approximated with a simple linear payment equation, if desired. However, in addition to the large payment reductions justified by the LCC analysis, this would also produce bonus payment factors that would be quite large. While bonus provisions are now widely used by agencies throughout the United States, top management has usually insisted that they be limited at some reasonable level, partly due to budget limitations and partly due to the possibility that a pavement might fail to achieve the expected extended life due to some condition not accounted for by the acceptance procedure.

A very practical way to address this issue, which has been used by one agency and has been well received by the construction industry in that State, is to use a compound payment equation as shown in figure 69. This has the twofold benefit of keeping the bonus provision within sensible limits and also dealing less harshly with a contractor whose work deviates only marginally from the desired quality level. It does, however, retain the safeguard of assessing large payment reductions for work that departs substantially from the desired level.

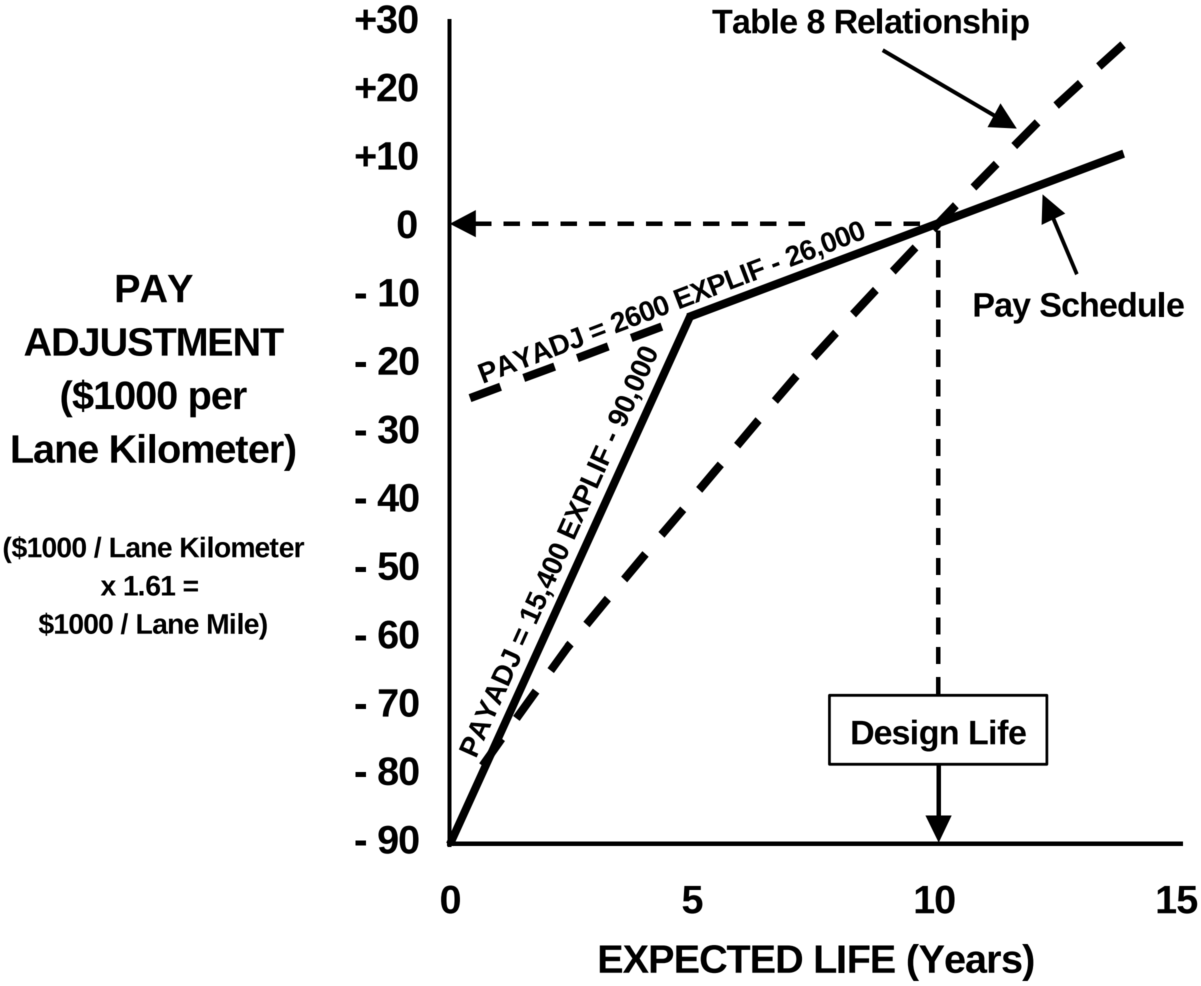


Figure 69. Payment Schedule Developed from Table 59

For this example, assume that management has decided that the maximum bonus to be paid will be no larger than $10,000 per lane kilometer. Since the maximum value of expected life returned by the performance model given by equation 115 is 13.8 years, and the design life used in this example for a typical overlay is 10 years, the upper portion of the payment equation must pass through the two points, $x = 13.8$, $y = 10{,}000$ and $x = 10$, $y = 0$. The slope, therefore, is computed as (10,000 – 0)/(13.8 – 10) = 2632, so 2600 will be used. The intercept is found by computing

0 – 2600(10 – 0) = –26,000 $/lane kilometer. The payment equation for this portion of the payment schedule is given by equation 117 and is also shown in figure 69.

$$PAYADJ = 2600\ (EXPLIF) - 26{,}000 \tag{117}$$

where: $PAYADJ$ = payment adjustment in units of $/lane kilometer.
$EXPLIF$ = expected life (years) obtained from equation 115.

For the lower part of the compound payment equation, it must be decided where the breakpoint is to be placed, and 5.0 years will be used for this example. The graph in figure 69 is used to determine –90,000 to be a suitable intercept. Then, since this payment equation must intersect the upper payment equation at $x = 5$, equation 117 is used to compute the ordinate at that point to be $y = 2600(5) - 26{,}000 = -13{,}000$. The slope is then obtained by computing (90,000 – 13,000)/(5 – 0) = 15,400, and the resulting payment equation is given by equation 118 and also shown in figure 69, in which all terms are as previously defined.

$$PAYADJ = 15{,}400\ (EXPLIF) - 90{,}000 \tag{118}$$

It should be noted that equations 117 and 118 represent just one of many suitable payment schedules that could be developed from the information in table 59 and the plot in figure 69. Either payment equation could be slightly steeper or shallower, provided they intersect at the breakpoint of *EXPLIF* = 5 years used in this example. The choice of breakpoint is purely a practical one, also, and a different breakpoint could have been used, if desired.

Finally, it is likely that most agencies would want to define an RQL in terms of expected life below which the agency has the option to require removal and replacement of the pavement at the contractor's expense. One possibility, suggested by the graph in figure 69, would be to define the RQL as a pavement whose expected life is less than 5.0 years. In this case, the lower portion of the payment schedule given by equation 118 would only come into play if the agency chose to waive the RQL provision.

Yet another possibility that may be appropriate is for the agency to specify that retests be performed to confirm the RQL condition before imposing the requirement to remove and replace the pavement. In this case, the upper payment equation would apply if the agency elected not to retest, and the lower payment equation would apply if the agency performed the retests, confirmed the RQL condition, but chose to waive the option to require removal and replacement.

Summary and Conclusions

A procedure has been presented that enables highway engineers with only a basic knowledge of engineering mathematics to use empirical construction data to develop realistic models for multiple quality characteristics, and to use those models to establish practical, effective, and defensible payment equations for QA specifications. The method is specifically designed to be easy to apply, and to avoid some of the problems to which other modeling methods may be prone, such as excessive complexity and the tendency to return unrealistic values when very large or very small input values must be used. A complete example was included for which performance data for three characteristics of HMAC pavement—in–place air voids, thickness, and smoothness—were used to develop a model for expected life. A simple LCC analysis was then applied to determine an appropriate payment schedule. The fact that this approach operates in two stages—first estimating expected life and then determining an appropriate payment adjustment—is believed to be desirable in that it will provide the type of information necessary to develop more accurate models in the future. Although the need to handle more than three quality characteristics at a time may be rare, the modeling method is sufficiently straightforward that additional characteristics can easily be accommodated, if necessary. Earlier versions of this approach have been used by one agency for several years, and their success has been reflected both in the quality achieved and the generally good working relationship the agency continues to have with the construction industry in the State.

APPENDIX M

OCPLOT Computer Program for Developing OC and EP Curves

Introduction

This appendix includes a brief explanation and discussion of the OCPLOT computer simulation program for developing OC and EP curves. This program was developed as part of FHWA Demonstration Project 89, and a much more thorough discussion of the OCPLOT program is provided in the manual for that project. (18)

In the case of pass/fail acceptance procedures, OC curves can be computed directly or constructed with the aid of specialized mathematical tables. For acceptance procedures with adjusted payment schedules, the construction of EP curves usually requires computer assistance. Program OCPLOT uses computer simulation to develop both OC and EP curves.

Features of Program OCPLOT

Program OCPLOT is designed to analyze the types of acceptance procedures typically used in the highway field. This includes pass/fail procedures, leading to the type of OC curve shown in figure M–1, and payment adjustment procedures, leading to the type of EP curve shown in figure M–2.

Figure M–3 lists some of the options that may be selected from the primary menu. The various items appear on the screen one at a time in a logical sequence, and later items are dependent upon the responses to earlier ones. The versatility of the program is apparent from the many different ways these selections might be combined.

When the selections from the first menu are complete, the menu will appear similar to that shown in figure M–4. The prompt “Press any key to continue” at the bottom of the display provides a pause that gives the user the opportunity to use the <ESC> key to go back and change some values or press the <PrintScreen> key to save the screen to a file if a record of the menu selections is desired. Striking almost any other key will cause the second menu in figure M–4 to appear.

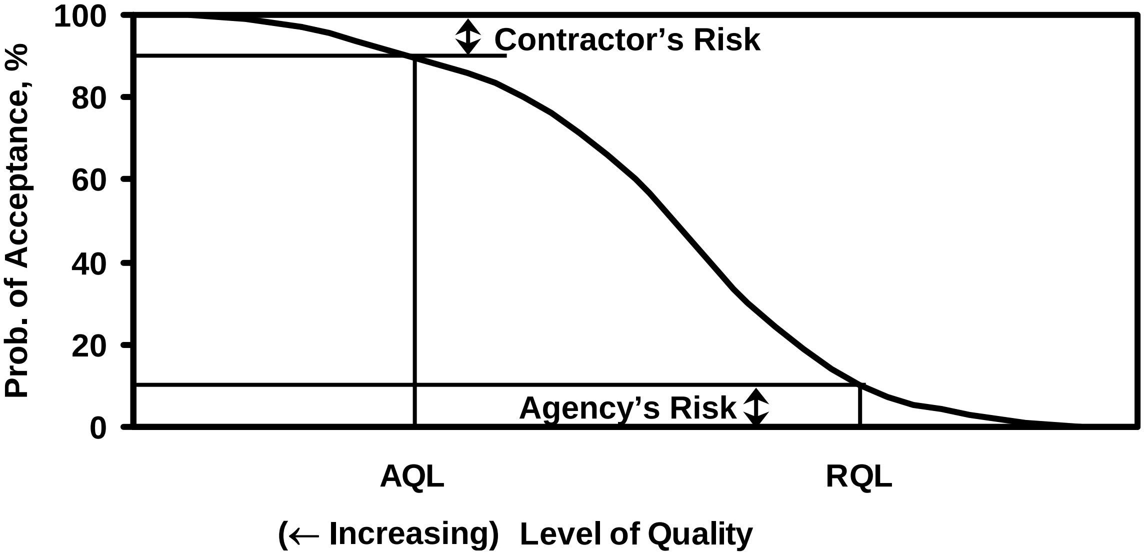


Figure M–1. Conventional OC Curve for Pass/Fail Acceptance Procedure

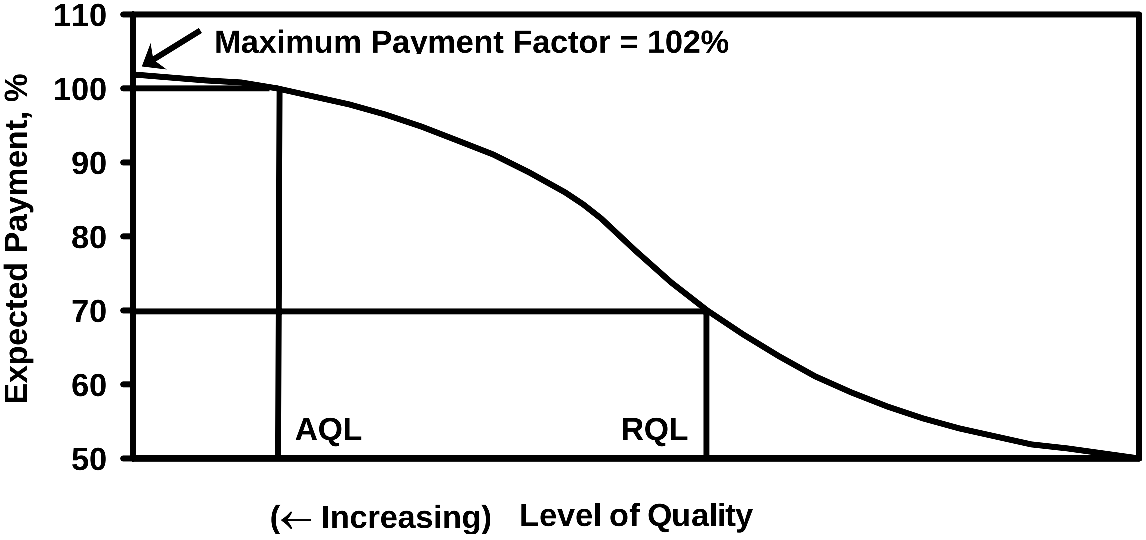


Figure M–2. Typical EP Curve for Acceptance Procedure with Adjusted Payment Schedule

ACCEPTANCE METHOD
Pass/Fail .... TYPE OF PLAN
Pay Adjustment
Attributes
Variables
QUALITY MEASURE
Percent Defective (PD)
Percent Within Limits (PWL)

LIMIT TYPE
Single-Sided
Double-Sided

PAY EQUATION TYPE
Linear/Nonlinear .... Enter Values

MAXIMUM PAY FACTOR
Yes .... Enter Values
No

ACCEPTABLE QUALITY LEVEL (AQL)
Enter Value

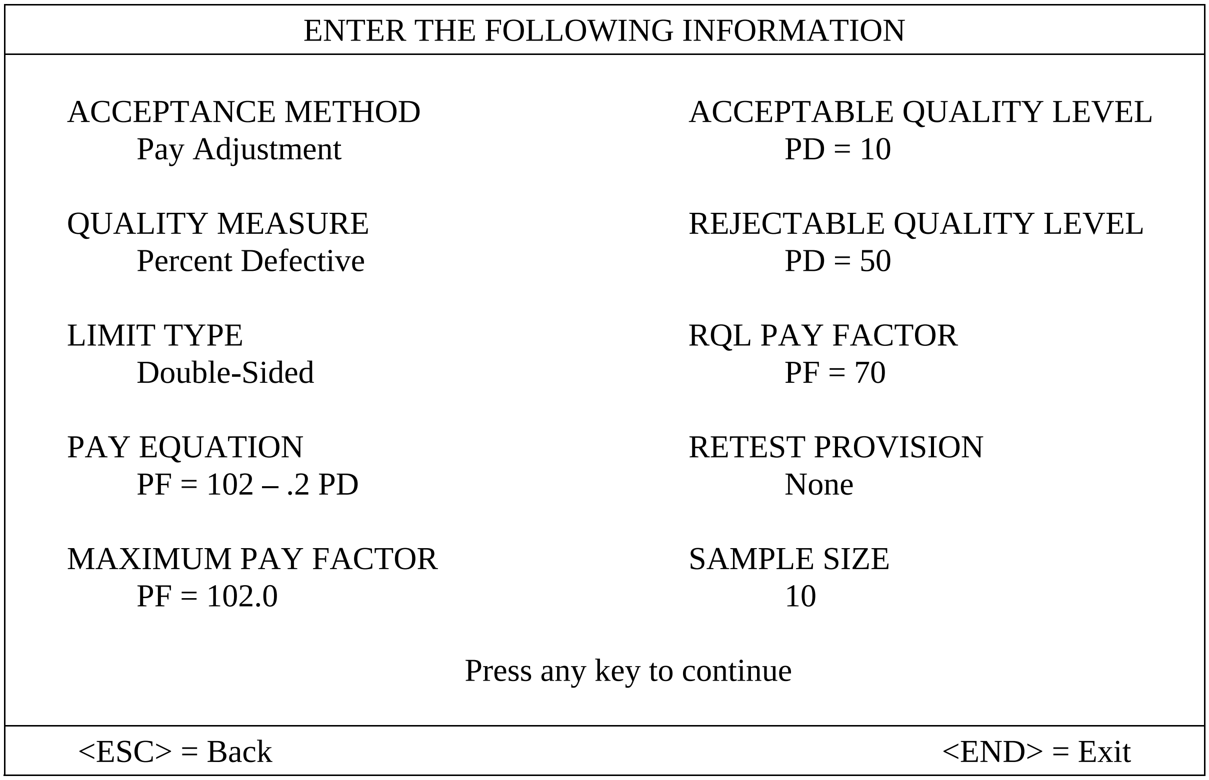
REJECTABLE QUALITY LEVEL (RQL)
Enter Value

RQL PROVISION
Yes .... Enter RQL Payment Factor
No

RETEST PROVISION
Yes .... INITIAL TESTS
No
Combined
Discarded
SAMPLE SIZE
Enter Value(s)

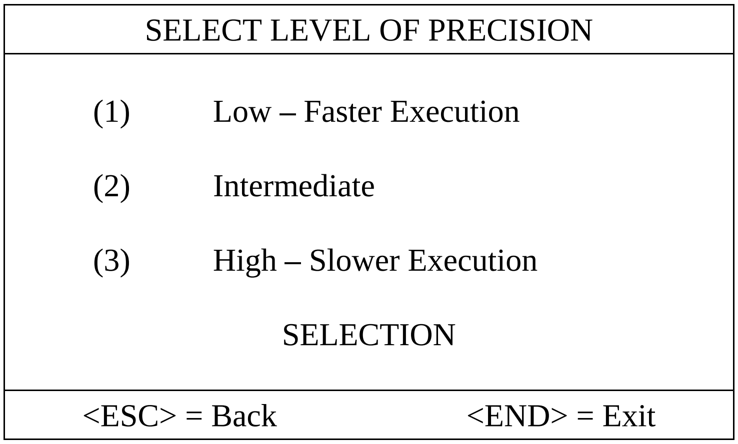
Figure M–3. Selections Available in Program OCPLOT

First Menu:



```
                    ENTER THE FOLLOWING INFORMATION

   ACCEPTANCE METHOD                    ACCEPTABLE QUALITY LEVEL
      Pay Adjustment                          PD = 10

   QUALITY MEASURE                      REJECTABLE QUALITY LEVEL
      Percent Defective                       PD = 50

   LIMIT TYPE                           RQL PAY FACTOR
      Double-Sided                            PF = 70

   PAY EQUATION                         RETEST PROVISION
      PF = 102 – .2 PD                        None

   MAXIMUM PAY FACTOR                   SAMPLE SIZE
      PF = 102.0                              10

                         Press any key to continue

   <ESC> = Back                                         <END> = Exit
```

Second Menu:



```
          SELECT LEVEL OF PRECISION

     (1)        Low – Faster Execution

     (2)        Intermediate

     (3)        High – Slower Execution

                    SELECTION

     <ESC> = Back                <END> = Exit
```

Figure M–4. First and Second Menus for Program OCPLOT

Because program OCPLOT uses computer simulation to analyze whatever acceptance procedure has been specified, it is very computationally intensive and the execution speed is dependent upon the level of precision selected from the second menu. Table M–1 lists the number of replications performed for the different levels of precision.

Table M–1. Program OCPLOT Precision Levels

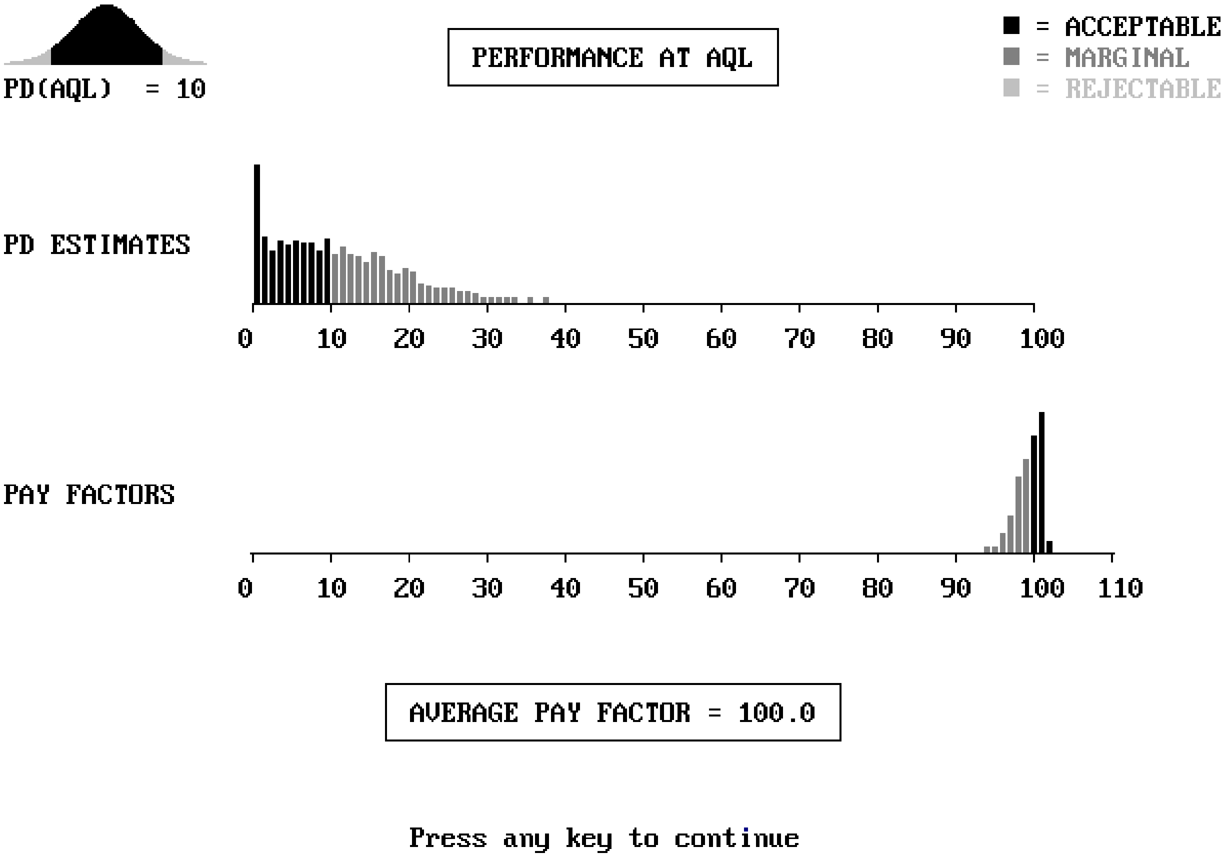
Precision Level	Number of Replications
1	200
2	1,000
3	5,000

Selection 1 provides the fastest execution, which is useful for exploratory work but may not be good enough to report as a final result. When this level is selected, 200 sample sets of the desired size are randomly generated from a normal population for each of several known levels of quality.

Each sample set is evaluated in accordance with the acceptance plan specified in the primary menu and the results are stored in memory for subsequent analysis. This is far more thorough and many times faster than testing the acceptance procedure with a field trial. (A field trial would be appropriate only if the procedure survives this initial check.)

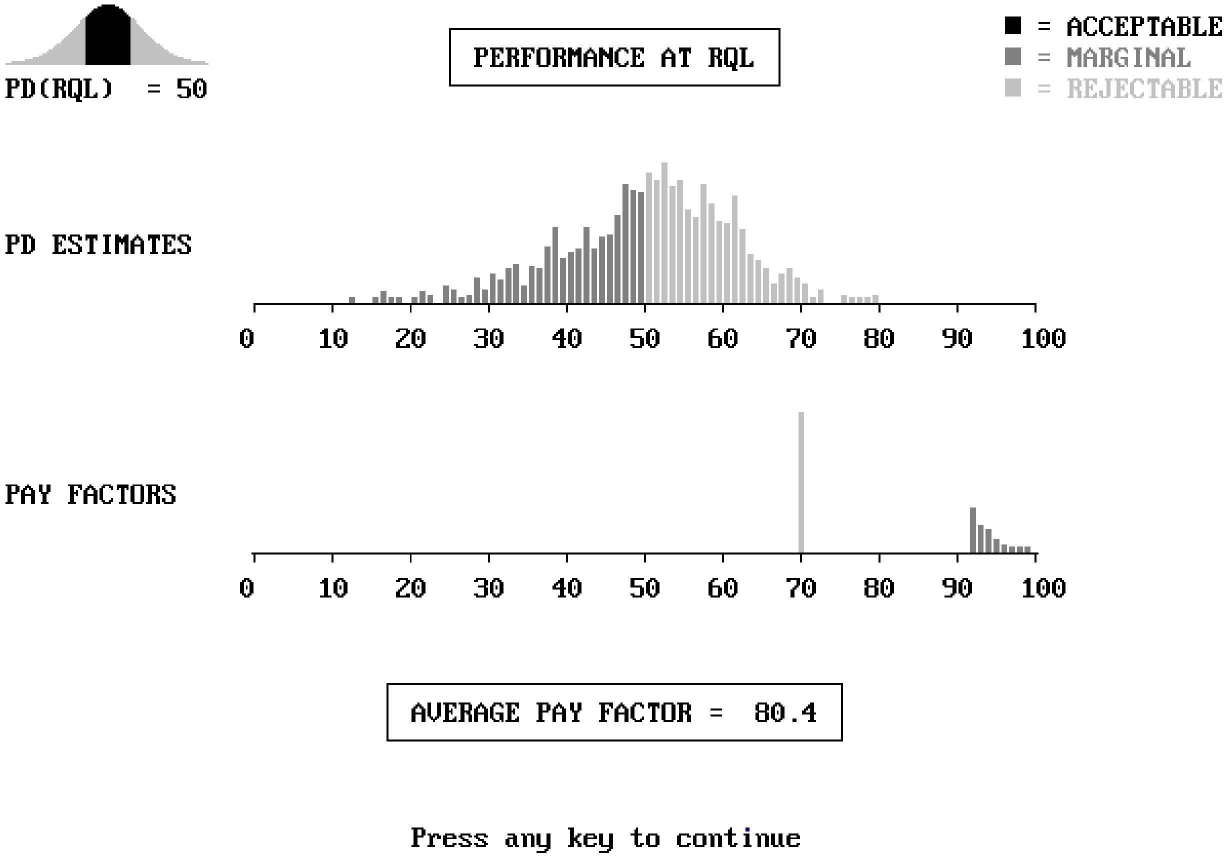
Selection 2 from the second menu provides an intermediate level of precision for which 1000 sample sets are generated at each quality level. This level of precision is usually satisfactory to report as a final result, producing points for the OC curve, representing probability of acceptance, or the EP curve, representing the expected payment factor, that are typically accurate to within one or two units. If still better precision is desired, selection 3 will cause 5000 sample sets to be generated at each quality level. This level of precision tends to produce a smooth line when the OC or EP curve is plotted.

Once the precision level is selected from the second menu, the computational process begins. For either low or intermediate precision, program OCPLOT displays detailed information at the two key points at which risk levels are usually expressed—the AQL and the RQL—as shown in figures M–5 and M–6. This serves two important purposes. For users less familiar with statistical estimation procedures and acceptance plans, the graphical displays at the AQL and RQL are both informative and educational. It may come as a surprise to some, for example, how widely distributed the quality estimates are, especially for small sample sizes. For users more familiar with statistical acceptance procedures, these displays provide assurance that the simulation process is working properly. The actual displays on a color monitor are color–coded to clearly distinguish acceptable and rejectable results and the corresponding payment factors.



```
PERFORMANCE AT AQL

AVERAGE PAY FACTOR = 100.0

Press any key to continue
```

Figure M–5. Typical Display at AQL at Intermediate Precision by Program OCPLOT



```
AVERAGE PAY FACTOR =  80.4

Press any key to continue
```

Figure M–6. Typical Display at RQL at Intermediate Precision by Program OCPLOT

Although the AQL and the RQL are probably the most important points at which it is desirable to know how the acceptance procedure will perform, it usually is useful to have a plot of the entire OC or EP curve that provides a picture of the performance over the complete range of quality that might be encountered. The prompt at the bottom of the screen instructs the user to strike any key to continue with this step to obtain the display shown in figure M–7. The horizontal and vertical axes and the two previously calculated points at the AQL and the RQL appear on the screen immediately. The remaining points appear one at a time at a speed determined by the level of precision that has been selected and the speed of the machine running the program.

After all the points have been calculated and plotted, the user may strike any key to connect the points with a solid line. Following this, the next key stroke will add vertical and horizontal lines highlighting the performance of the acceptance plan at the AQL and RQL, as shown in figure M–8. And, like the histograms in figures M–5 and M–6, any of these displays may be saved to a file by using the <PrintScreen> key.

Following this display, striking a key will produce the menu shown in figure M–9. If the first item in this menu is selected, the output shown in figure M–10 is displayed. This permits the user to save the values of the data points shown in figure M–7 from which the OC curve was constructed. The other selections in this menu make it possible to return to earlier points in the input stage of the program or to exit the program.

Example M–1: Pass/Fail Attributes Procedure

Attributes acceptance procedures are based on measures that are counted rather than computed, such as the number of defects on an item of production or the number of test results falling outside specification limits. In contrast, variables acceptance procedures are based on statistical parameters that are computed, such as the mean and standard deviation, and lead to the estimation of PD or PWL.

Advantages of attributes procedures are that they are simple to apply and they require no assumptions about the underlying distributional form of the population being sampled. For example, a typical attributes procedure might require that a sample of size N = l0 be taken and that no more than C = 2 test results may be outside the specification limits, where “C” is referred to as the acceptance number. A disadvantage is that they require larger sample sizes to achieve the equivalent discriminating power of a variables procedure. Variables procedures, because of their inherently greater efficiency, are generally preferred whenever the requirement for a normally–distributed population is reasonably satisfied.

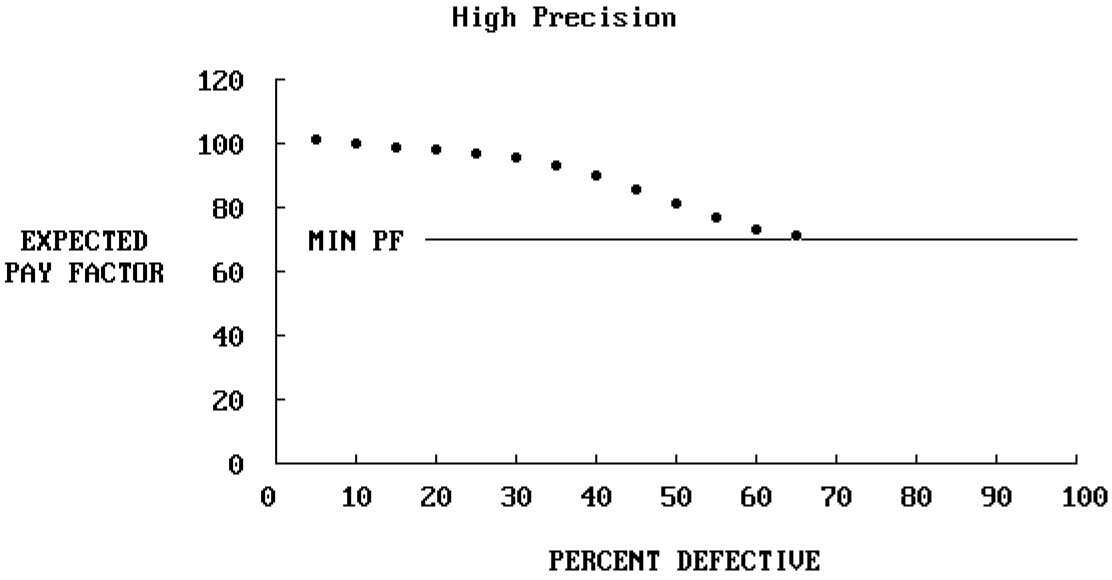


Figure M–7. Points on EP Curve Plotted by Program OCPLOT

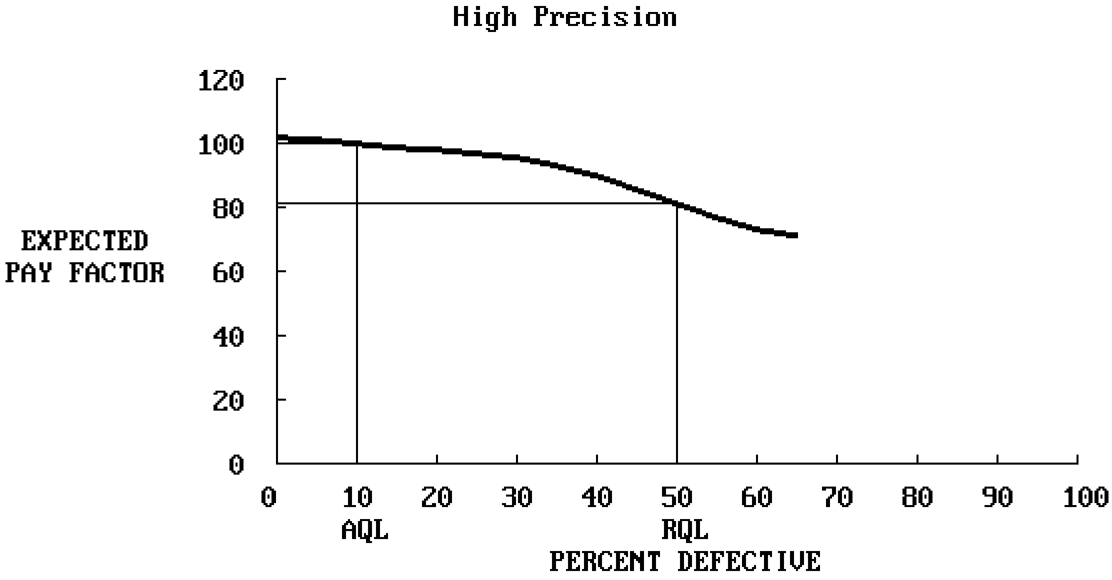
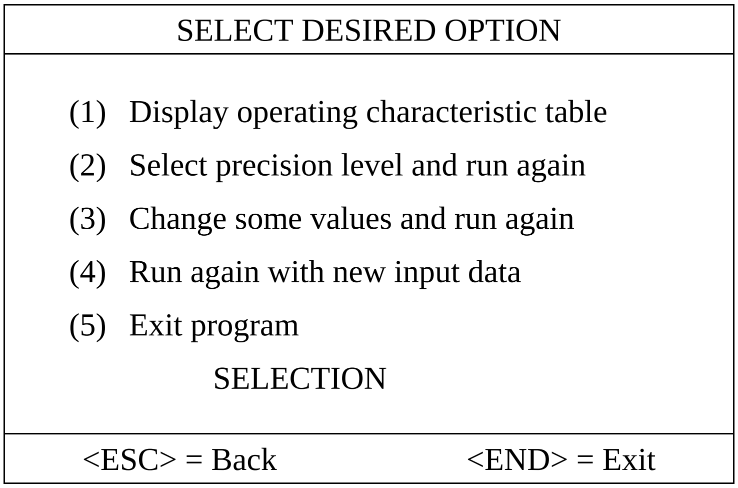


Figure M–8. Display of EP Curve Plotted by Program OCPLOT with AQL and RQL Performance Highlighted



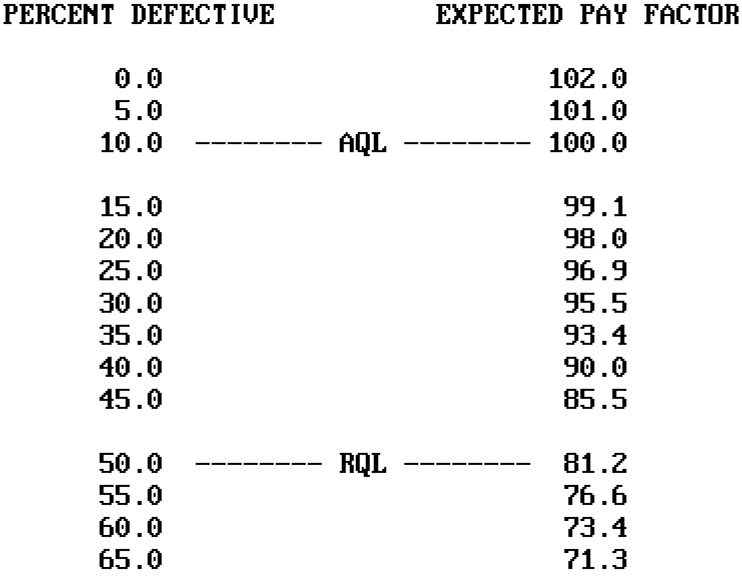
SELECT DESIRED OPTION

(1) Display operating characteristic table
(2) Select precision level and run again
(3) Change some values and run again
(4) Run again with new input data
(5) Exit program

SELECTION

<ESC> = Back <END> = Exit

Figure M–9. Third Menu for Program OCPLOT



```
PERCENT DEFECTIVE          EXPECTED PAY FACTOR

      0.0                         102.0
      5.0                         101.0
     10.0  -------- AQL -------- 100.0

     15.0                          99.1
     20.0                          98.0
     25.0                          96.9
     30.0                          95.5
     35.0                          93.4
     40.0                          90.0
     45.0                          85.5

     50.0  -------- RQL --------   81.2
     55.0                          76.6
     60.0                          73.4
     65.0                          71.3
```

Figure M–10. Display of Numerical Values of Points on EP Curve Computed at High Precision by Program OCPLOT

Although the vast majority of highway construction measures tend to be normally distributed, there are some that are not. A physical constraint close to the desired target value often produces non–normality. For example, depth of cover over the top mat of reinforcing steel in a bridge deck may tend to be skewed because it is physically impossible (provided the steel is embedded) for the cover to be less than zero but there is less of a restriction on how deep the mat might be. Similarly, if the target level for percent air voids in bituminous pavement is fairly low, the physical limit of zero may tend to skew this distribution toward larger values. Another condition that can produce non–normality is the combining in the same lot of two distinctly different populations. As a general rule, a conscious effort should be made to avoid this condition when applying a statistical specification.

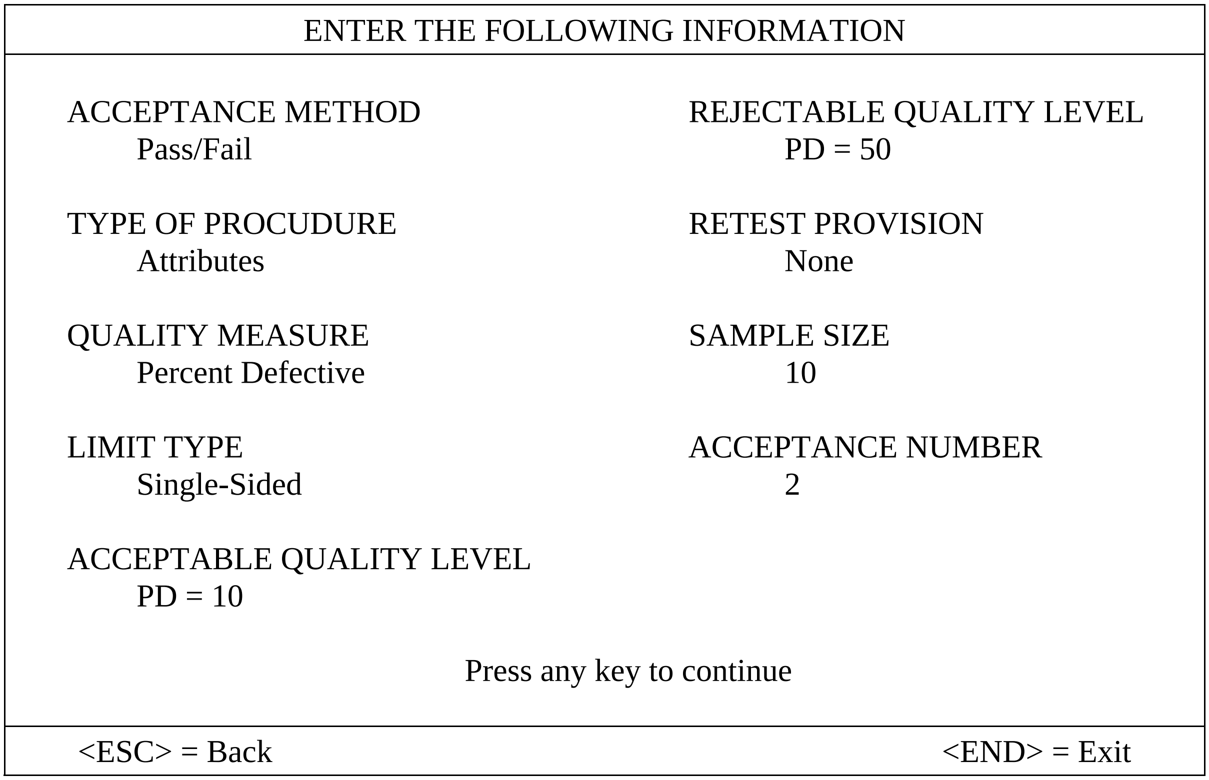
Because situations may arise in highway construction that warrant their use, program OCPLOT provides the capability of analyzing attributes acceptance procedures. Although it would be possible to develop a payment schedule based on an attributes procedure, their use has been limited almost exclusively to pass/fail applications. The following is presented as a generic example of such a procedure.

It is assumed for this example that the statistical quality measure is the percent defective (PD), the percentage of the lot falling outside specification limits. The acceptable quality level (AQL) and the rejectable quality level (RQL) are defined as PD = 10 and PD = 50, respectively. It is desired to develop an acceptance procedure that will balance the risks at 0.05. In other words, if the contractor provides work that is exactly at the AQL, there is to be a 0.05 chance that it will erroneously be rejected and, at the other extreme, if the work is truly at the RQL, there is to be a 0.05 chance that it will erroneously be accepted.

An acceptance plan is required, stated in terms of the sample size (N) and the acceptance number (C) that will produce the desired risks. For both risks to be 0.05, an OC curve is desired that indicates probabilities of acceptance at the AQL and the RQL of 0.95 and 0.05, respectively. Because both N and C must be integer values, the resultant risk levels vary in discrete steps, and it usually is not possible to obtain exactly the desired risk values. To find a plan that matches the desired risk levels as closely as possible, it is necessary to examine several plans.

This can easily be accomplished with program OCPLOT by selecting “Pass/Fail” and “Attributes” from the opening menu, followed by other appropriate selections and ending with a trial combination of N and C. It is usually advantageous to make the first few runs at low precision (selected from the second menu) in order to speed up the trial–and–error process.

Figure M–11 shows the completed menu for the initial try with N = 10 and C = 2 and figure M–12 shows the resulting acceptance probabilities obtained at high precision. It can be seen in figure M–12 that the probability of acceptance of 0.052 at the RQL is close to the desired value, while the probability of acceptance of 0.927 at the AQL is too low.



ENTER THE FOLLOWING INFORMATION

ACCEPTANCE METHOD
Pass/Fail

TYPE OF PROCUDURE
Attributes

QUALITY MEASURE
Percent Defective

LIMIT TYPE
Single-Sided

ACCEPTABLE QUALITY LEVEL
PD = 10

REJECTABLE QUALITY LEVEL
PD = 50

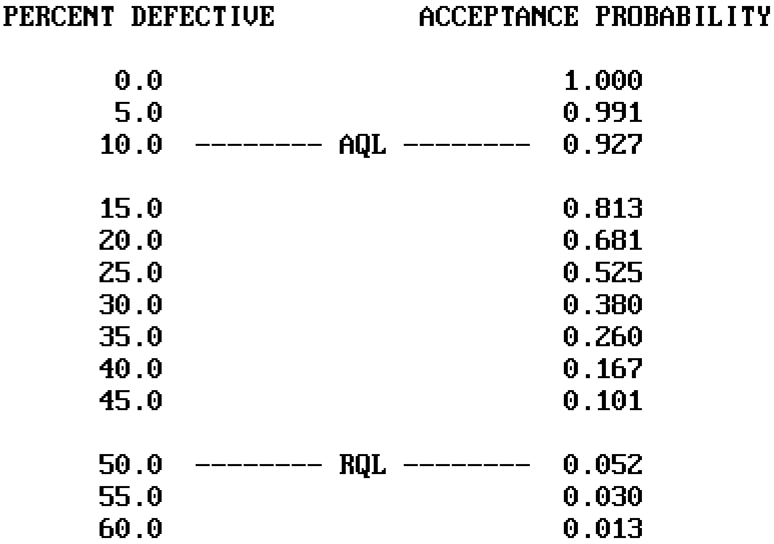
RETEST PROVISION
None

SAMPLE SIZE
10

ACCEPTANCE NUMBER
2

Press any key to continue

Figure M–11. Completed Menu for Analysis of Pass/Fail Attributes Acceptance Plan with N = 10 and C = 2



```
PERCENT DEFECTIVE        ACCEPTANCE PROBABILITY

       0.0                          1.000
       5.0                          0.991
      10.0  -------- AQL --------   0.927

      15.0                          0.813
      20.0                          0.681
      25.0                          0.525
      30.0                          0.380
      35.0                          0.260
      40.0                          0.167
      45.0                          0.101

      50.0  -------- RQL --------   0.052
      55.0                          0.030
      60.0                          0.013
```

Figure M–12. Numerical Values of Points on EP Curve for Pass/Fail Attributes Acceptance Plan with N = 10 and C = 2

The results of several attempts are presented in table M–2. The values in this table were all obtained at high precision and here it can be seen that the plan having N = 13 and C = 3 comes the closest to meeting the desired acceptance probabilities of Prob. ≥ 0.95 and Prob. ≤ 0.05 at the AQL and RQL, respectively. This is a trial and error process, but, with a little experience, it usually is possible to find a suitable plan relatively quickly.

Table M–2. Performance of Attributes Acceptance Plans

Percent Defective (PD)	**Probability of Acceptance**			
	N = 10 C = 2	**N = 12 C = 3**	**N = 13 C = 3**	**N = 14 C = 3**
0	1.000	1.000	1.000	1.000
10 (AQL)	0.927	0.975	0.966	0.960
20	0.681	0.810	0.747	0.698
30	0.380	0.491	0.430	0.350
40	0.167	0.226	0.163	0.130
50 (RQL)	0.052	0.079	0.048	0.028
60	0.013	0.014	0.006	0.011

The attributes acceptance plan meeting the requirements of this example requires that a total of N = 13 test specimens be taken from random locations within each lot and that, after the appropriate tests have been performed, no more than C = 3 of the test results shall be outside specification limits. The use of such a plan ensures that truly AQL work will be accepted about 95 percent of the time and that truly RQL work will be accepted only about 5 percent of the time.

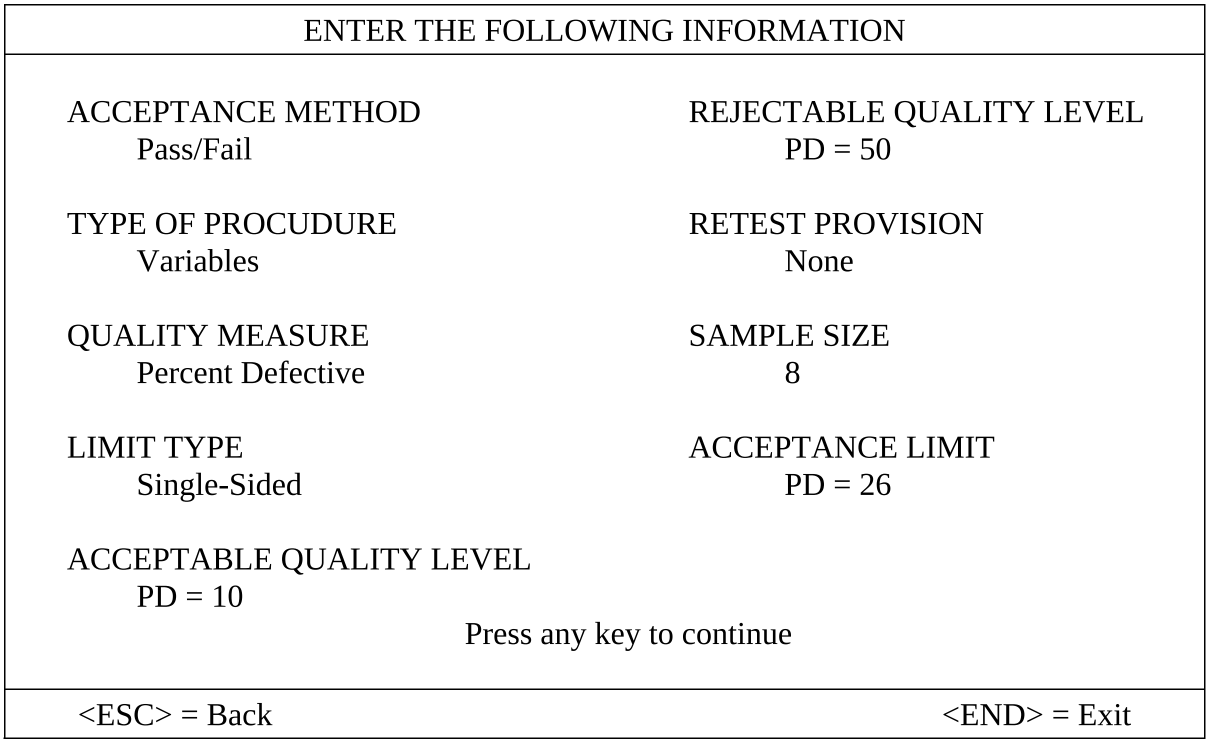
Example M–2: Pass/Fail Variables Procedure

Variables procedures are based on the assumption that the population being sampled is at least approximately normally distributed. They involve the computation of statistical parameters, such as the mean and standard deviation, to estimate PD or PWL. Provided the normality assumption is sufficiently well satisfied, variables procedures can provide essentially the same discriminating power as equivalent attributes plans, but with a substantially smaller sample size.

To illustrate this last statement, a variables procedure is sought that will have essentially the same discriminating power (OC curve) as the attributes procedure developed in example M–1. Like the previous example, this involves a trial and error process with which program OCPLOT can be extremely helpful. For this example, the selections “Pass/Fail” and “Variables” are made from the opening menu, followed by other appropriate selections and ending with trial values for the sample size and acceptance limit.

This trial–and–error process proceeds more quickly if a low level of precision is selected for the initial attempts. When it appears that the appropriate combination of sample size and acceptance limit has been found, this should be checked at intermediate precision. Further minor adjustments may then be necessary before confirming the result at high precision.

Figure M–13 shows the completed menu for the variables acceptance plan that was ultimately selected. The actual numerical values on the OC curve are shown in table M–3, which summarizes the results of examples M–1 and M–2. In this table it can be seen that the performance of the variables plan very closely matches that of the attributes plan in example M–1 (table M–2, N = 13, C = 3). The dramatic difference is that this has been accomplished with a sample size of N = 8, whereas the attributes plan required a sample size of N = 13. In this case, which assumes that the normality assumption is satisfied, the use of a variables plan results in a direct savings in sampling and testing costs of nearly 40 percent.



ENTER THE FOLLOWING INFORMATION

ACCEPTANCE METHOD
Pass/Fail

TYPE OF PROCUDURE
Variables

QUALITY MEASURE
Percent Defective

LIMIT TYPE
Single-Sided

ACCEPTABLE QUALITY LEVEL
PD = 10

REJECTABLE QUALITY LEVEL
PD = 50

RETEST PROVISION
None

SAMPLE SIZE
8

ACCEPTANCE LIMIT
PD = 26

Press any key to continue

Figure M–13. Completed Menu for Pass/Fail Variables Acceptance Plan Example

Table M–3. Comparison of Operating Characteristics of Attributes and Variables Acceptance Plans

Percent Defective	**Probability of Acceptance**	
	Attributes Plan (N =13, C= 3)	**Variables Plan (N = 8, Acc Limit = 26 PD)**
0	1.000	1.000
5	0.997	0.996
10 (AQL)	0.966	0.954
15	0.884	0.841
20	0.744	0.693
25	0.596	0.544
30	0.430	0.376
35	0.279	0.256
40	0.163	0.157
45	0.091	0.088
50 (RQL)	0.048	0.048
55	0.024	0.024
60	0.006	0.009

Example M–3: Analysis of Payment Equation

Although the pass/fail acceptance procedures discussed in examples M–1 and M–2 may be useful for highway construction applications, the use of acceptance procedures with adjusted payment schedules is generally of greater interest. The proper design of such plans is critical to their performance, and poorly conceived plans may be either totally ineffective or impractically severe. Neither problem may be apparent, however, until the plan has been analyzed by constructing the EP curve.

Equation M–1 gives a payment schedule proposed for use by an owner. The AQL was defined as PWL_{SPEC} = 90 and, since there was no incentive payment provision, the maximum payment factor was limited to 100 percent.

$$PR = \mathrm{PWL}_{SPEC} - \mathrm{PWL}_{COMP} \qquad \text{(M–1)}$$

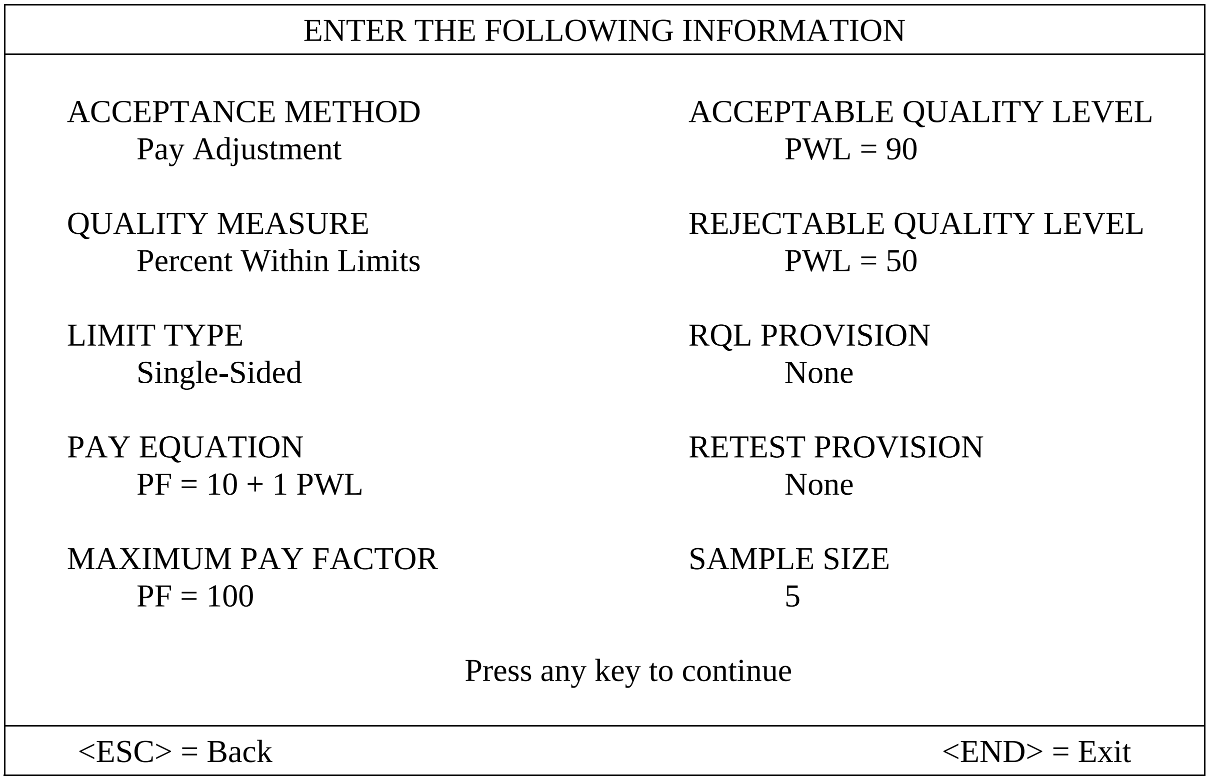
where: PR = payment reduction (percent).
PWL_{SPEC} = specified PWL at the AQL.
PWL_{COMP} = PWL computed from test values.

To transform this equation into a form that can be handled by program OCPLOT, it is necessary to express it in terms of the payment factor rather than the payment reduction. By substituting $\mathrm{PWL}_{SPEC} = 90$ and $PR = 100 - PF$ into equation M–1, equation M–2 is obtained, subject to the restriction that $PF \leq 100$.

$$PF = 10 + \mathrm{PWL}_{COMP} \qquad \text{(M–2)}$$

where: PF = payment factor (percent)
PWL_{COMP} = PWL computed from test values.

To judge the effectiveness of this payment equation, the EP curve will be developed for a typical sample size of N = 5. The completed input menu is shown in figure M–14. Following this, an intermediate level of precision was selected from the second menu to obtain the display at the AQL shown in figure M–15. Finally, the program was run at high precision, which produced the displays shown in figures M–16 and M–17.



```
                    ENTER THE FOLLOWING INFORMATION

  ACCEPTANCE METHOD                    ACCEPTABLE QUALITY LEVEL
       Pay Adjustment                          PWL = 90

  QUALITY MEASURE                      REJECTABLE QUALITY LEVEL
       Percent Within Limits                   PWL = 50

  LIMIT TYPE                           RQL PROVISION
       Single-Sided                            None

  PAY EQUATION                         RETEST PROVISION
       PF = 10 + 1 PWL                         None

  MAXIMUM PAY FACTOR                   SAMPLE SIZE
       PF = 100                                5

                       Press any key to continue

   <ESC> = Back                                    <END> = Exit
```

Figure M–14. Completed Menu for Analysis of Payment Equation

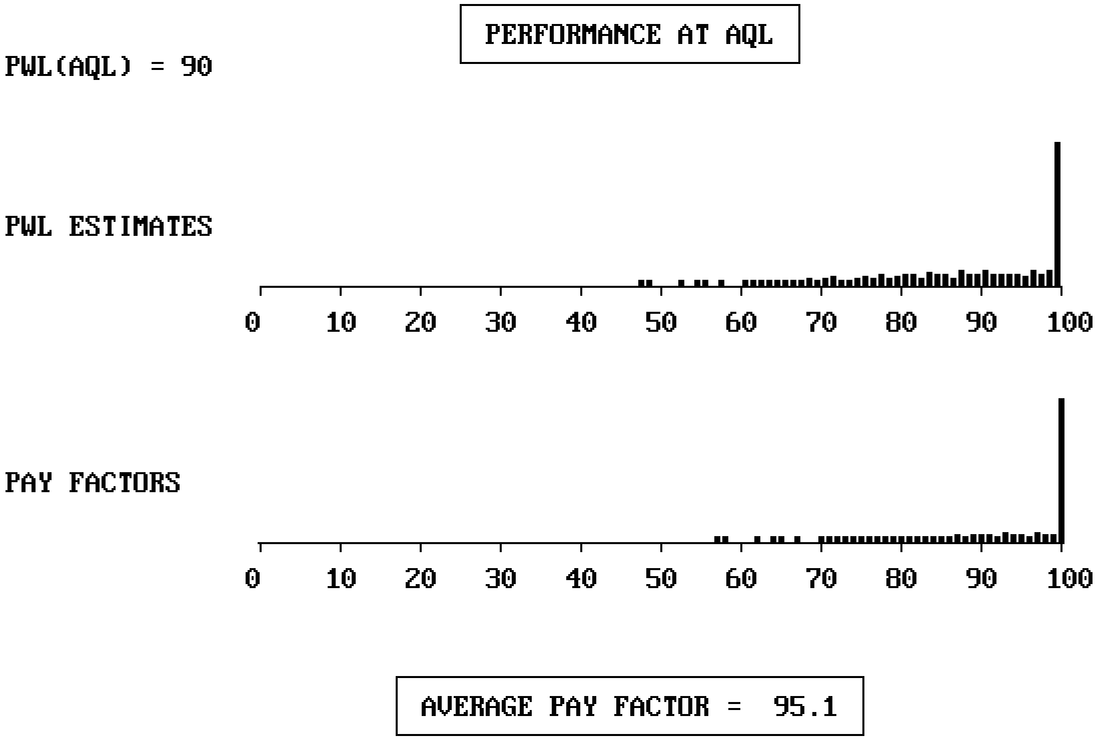


Figure M–15. Performance of Payment Equation at AQL

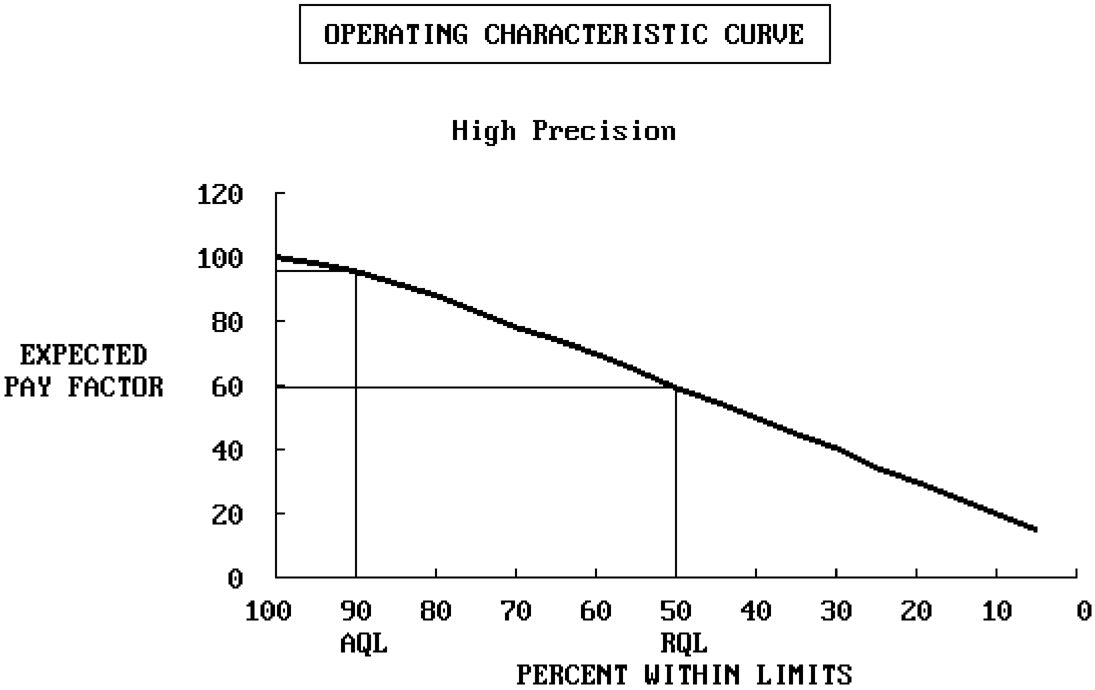
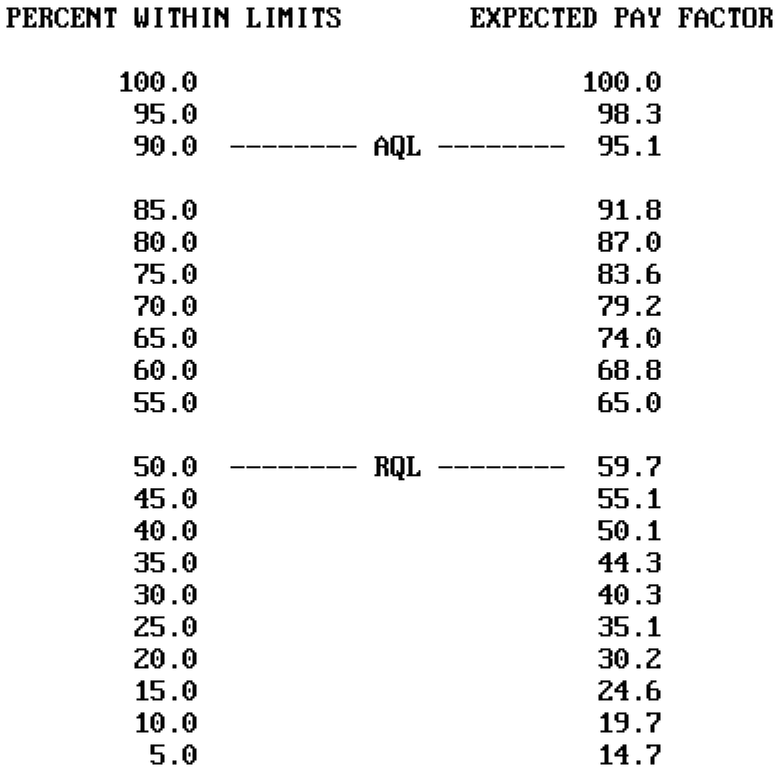


Figure M–16. EP Curve for Analysis of Payment Equation



```
PERCENT WITHIN LIMITS       EXPECTED PAY FACTOR

         100.0                      100.0
          95.0                       98.3
          90.0  -------- AQL --------  95.1

          85.0                       91.8
          80.0                       87.0
          75.0                       83.6
          70.0                       79.2
          65.0                       74.0
          60.0                       68.8
          55.0                       65.0

          50.0  -------- RQL --------  59.7
          45.0                       55.1
          40.0                       50.1
          35.0                       44.3
          30.0                       40.3
          25.0                       35.1
          20.0                       30.2
          15.0                       24.6
          10.0                       19.7
           5.0                       14.7
```

Figure M–17. Numerical Values for Points on EP Curve for Analysis of Payment Equation

It can be seen in figure M–15 that there is a serious problem with this acceptance procedure. A contractor who performs consistently at the AQL will receive an average payment reduction of nearly 5 percent. To emphasize the impact this would have on the construction industry, this means that a contractor responsible for $10 million worth of work under this specification over the course of a construction season would on the average be penalized approximately $500,000 for successfully providing the level of quality that was explicitly defined as acceptable in the contract documents.

This example illustrates the situation discussed previously, whereby the failure to include an incentive payment provision with this type of specification prevents the acceptance procedure from paying an average of 100 percent when the work is exactly at the AQL. The reason for this can be seen with the help of figure M–15. The upper histogram in this figure represents the distribution of 1000 PWL estimates, each obtained by randomly sampling a population that is exactly of AQL quality. These sample PWL estimates range from a low of about 47 percent to a high of 100 percent because of the inherent variability of the sampling process. However, the long–term average of these estimates will always be extremely close to the true value (the AQL in this case) because the PD/PWL estimation process is an unbiased statistical estimation procedure. The problem arises because the payment schedule does not permit this unbiased property to extend to the distribution of payment factors. Because there is no incentive payment provision, all PWL estimates that are greater than the true PWL of 90 receive the maximum payment factor of 100 percent, while all those below the true value receive some degree of payment reduction. Since approximately half the lots will receive payment reductions and the other half will receive 100 percent payment, the net result is that the average payment factor for AQL work will be substantially lower than 100 percent.

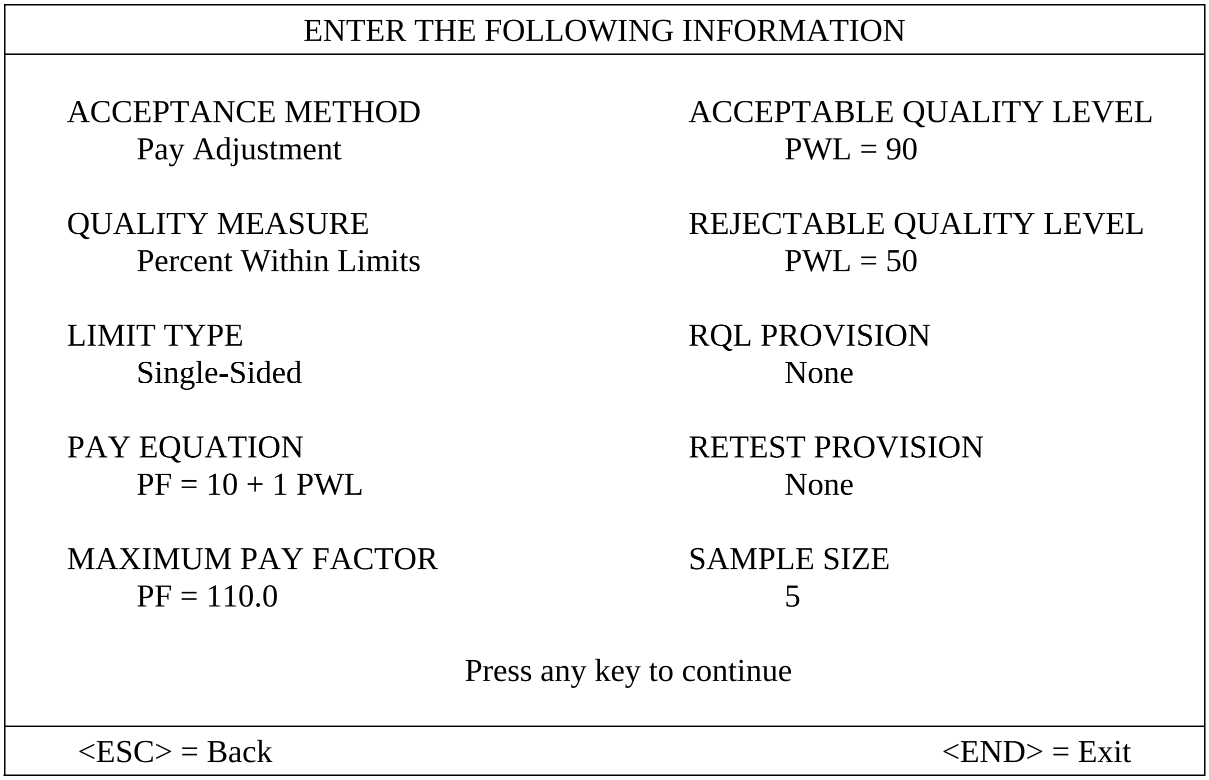
This situation is clearly misleading and unfair, yet it exists in current QA specifications. It is not difficult to correct, however, as demonstrated in the next example.

Example M–4: Effect of Incentive Payment Provision

The problem described in the preceding example, in which truly AQL work receives a substantial payment reduction, is easy to correct. All that is required is the inclusion of an incentive payment provision as part of the acceptance procedure. In equation M–2, this would mean removing the restriction that the maximum payment factor cannot exceed 100 percent.

Ordinarily, the maximum payment factor and the slope of the payment equation should be consistent with established (or estimated) performance relationships and the anticipated economic consequences of any departures from the specified AQL. To be consistent with this example, a maximum incentive payment factor of 110 percent will have to be used. In actual practice, however, most agencies have used incentive payment provisions of 105 percent or less.

To confirm that this will solve the problem, program OCPLOT was run an additional time with the identical input used for example M–3 except that when the prompt “MAXIMUM PAY FACTOR” appeared, selection 2 was chosen, indicating that no upper limit is placed on the payment equation. The program then automatically computed and displayed the maximum payment factor of $PF = 110.0$, as shown in the completed menu in figure M–18. The performance at the AQL is displayed in figure M–19, where it is seen that the expected payment factor is 100 percent, as desired.



ENTER THE FOLLOWING INFORMATION

ACCEPTANCE METHOD
Pay Adjustment

QUALITY MEASURE
Percent Within Limits

LIMIT TYPE
Single-Sided

PAY EQUATION
PF = 10 + 1 PWL

MAXIMUM PAY FACTOR
PF = 110.0

ACCEPTABLE QUALITY LEVEL
PWL = 90

REJECTABLE QUALITY LEVEL
PWL = 50

RQL PROVISION
None

RETEST PROVISION
None

SAMPLE SIZE
5

Press any key to continue

Figure M–18. Completed Menu for Analysis of Acceptance Procedure with Incentive Payment Provision

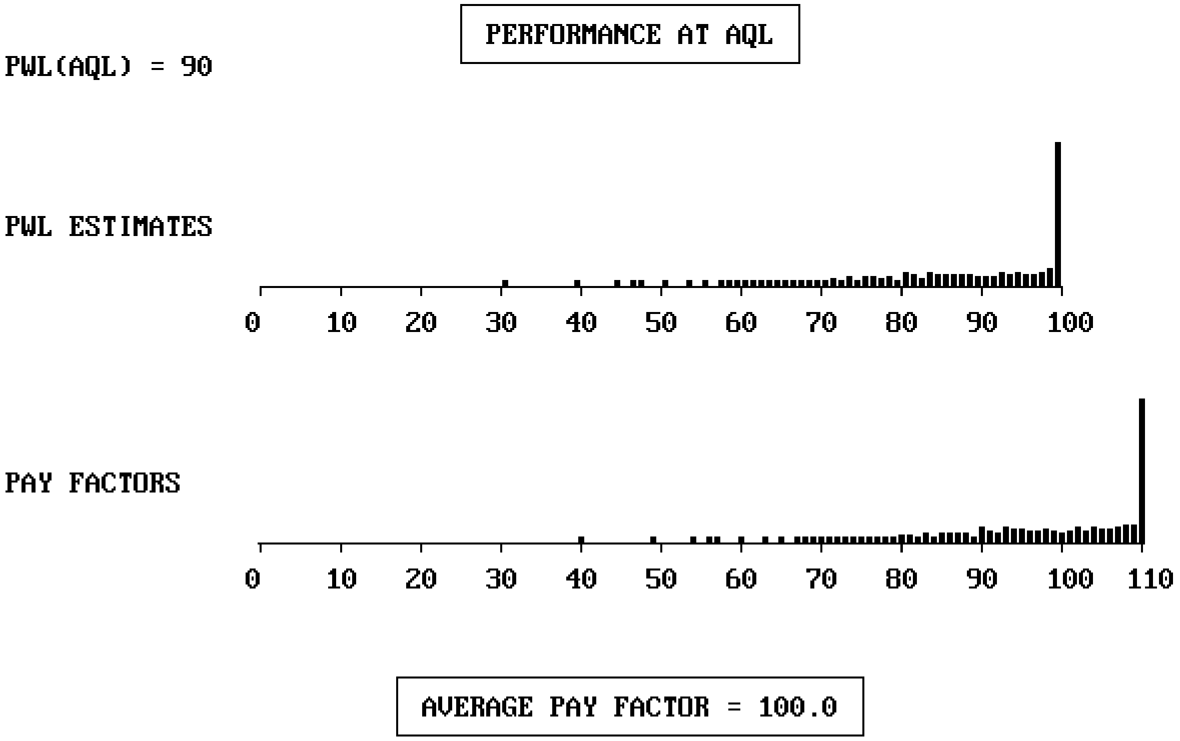


Figure M–19. Performance at AQL of Acceptance Procedure with Inventive Payment Provision

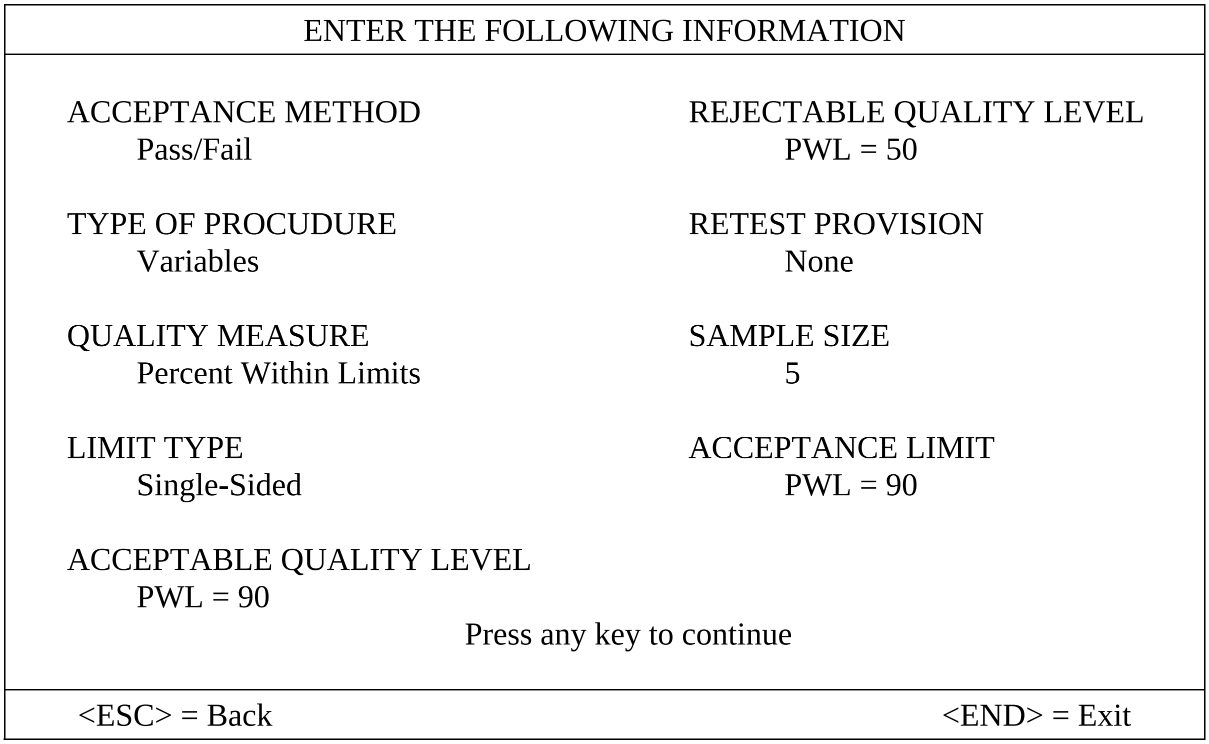
Example M–5: OC Curves for Payment Adjustment Plans

It was shown in chapter 7 that multiple OC curves could be plotted to illustrate the performance of acceptance plans with payment adjustment provisions. Each of these curves is associated not with the probability of acceptance, but with the probability of receiving greater than or equal to a given payment level. For example, suppose that equation M–3 is selected to determine the payment for a lot based on a sample of size N = 5.

$$PF = 55 + (0.5 \times \text{PWL}) \tag{M–3}$$

From this equation it can be seen that full payment, i.e., *PF* = 100 percent, is obtained for a computed lot PWL of 90. Lot PWL values greater than 90 will lead to an incentive payment, while lot PWL values less than 90 will lead to a disincentive payment.

The probability of receiving a payment greater than or equal to 100 percent therefore is equivalent to the probability of receiving an individual estimated lot PWL of 90 or greater. By using the Pass/Fail option and selecting 90 as the ACCEPTANCE LIMIT, the OCPLOT program can be used to determine this OC curve. The completed OCPLOT menu for this situation is shown in figure M–20, and the corresponding OC curve is shown in figure M–21.



ENTER THE FOLLOWING INFORMATION

ACCEPTANCE METHOD
Pass/Fail

TYPE OF PROCUDURE
Variables

QUALITY MEASURE
Percent Within Limits

LIMIT TYPE
Single-Sided

ACCEPTABLE QUALITY LEVEL
PWL = 90

REJECTABLE QUALITY LEVEL
PWL = 50

RETEST PROVISION
None

SAMPLE SIZE
5

ACCEPTANCE LIMIT
PWL = 90

Press any key to continue

<ESC> = Back
<END> = Exit

Figure M–20. Completed Menu for Using OCPLOT to Determine the Probability of Receiving Greater than or Equal to 100 Percent Payment for Example M–5

Similarly, the probability of receiving greater than or equal to 95 percent payment is equivalent to the probability of receiving an individual estimated lot PWL of 80 or greater (obtained by plugging *PF* = 90 into equation M–3 and solving for PWL). OCPLOT can once again be used with the Pass/Fail option and with 80 as the ACCEPTANCE LIMIT to determine this OC curve. This can be repeated for other payment levels to develop the family of OC curves shown in figure M–22.

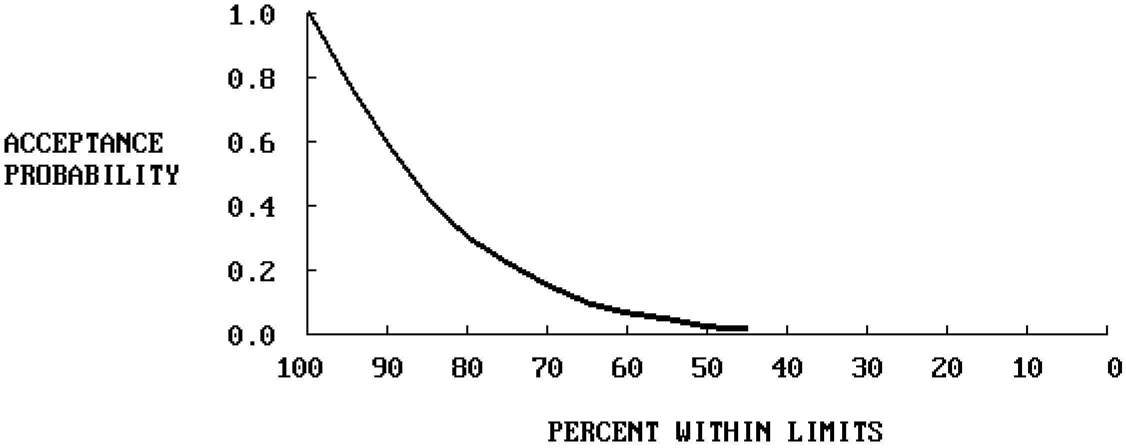


Figure M–21. OC Curve for the Probability of Receiving Greater than or Equal to 100 Percent Payment for Example M–5

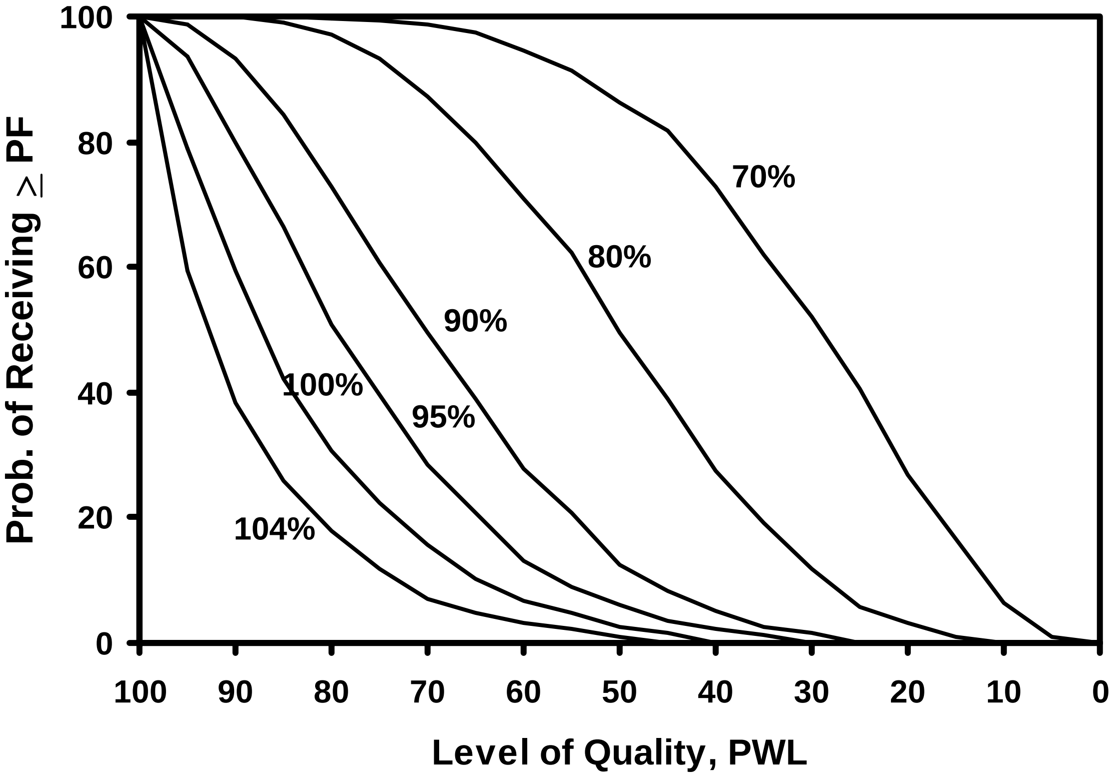


Figure M–22. Family of OC Curves Showing the Performance of the Acceptance Plan Using a Sample of Size N = 5 and Payment Equation M–3

APPENDIX N

Job Specifications for NJDOT Statistical Engineer Positions

NEW JERSEY DEPARTMENT OF CIVIL SERVICE

DIVISION OF CLASSIFICATION AND COMPENSATION

STATISTICAL ENGINEER I

DEFINITION

Under the direction of a Bureau Chief, Division of Research and Demonstration, Department of Transportation, supervises the performance of work involving the application of complex statistical and mathematical techniques to the solution of engineering and other problems; does related work as required.

EXAMPLES OF WORK

Acts as consultant on statistical matters related primarily to engineering applications for all Departmental units desiring this service.

Meets with various bureau heads to assess how statistical techniques can most effectively be applied to solve various engineering and other problems.

Supervises and instructs subordinates (his own and those of others) in the performance of various statistical analyses.

Is solely responsible and accountable for recommendations made by the Statistical Engineering Group. Supervision received will be broad and general, not technical.

Analyzes and prepares statistical evidence for legal proceedings such as the defense of lawsuits against the Department concerning construction specifications and the initiation of civil suits by the Department to reclaim damages for inferior construction work. Provides expert testimony, as required.

Prepares and writes statistically-oriented engineering specifications, establishing reasonable and balanced risks to promote smooth implementation and acceptance and reduce the likelihood of subsequent litigation.

Uses computer simulation to test statistical and other engineering specifications to assure that they are both fair and effective.

Analyzes engineering tolerances required for various materials and construction specifications to assure that they are realistic from the standpoint of being obtainable, and effective from the standpoint of accomplishing the desired engineering function.

Investigates various engineering applications of statistical analysis to assure that underlying theoretical assumptions are sufficiently satisfied.

Plans statistical studies including the experimental design, type and quantity of data to be collected, and methods of analysis.

Plans correlation and calibration studies of various engineering tests and measurement techniques and develops precision statements. Makes recommendations as to which methods are superior taking both statistical and engineering factors into account.

Calculates confidence limits on various types of data and information obtained by the Department to provide a measure of its significance for decision-making purposes. This may include environmental analysis parameters, traffic and accident data, and maintenance costs, among others,

Uses operations research (linear programming) techniques or other appropriate statistical or engineering-economics methods to solve optimization problems such as determining where limited appropriations may most effectively be spent.

Performs hypothesis tests to determine whether a variety of experimental features in equipment or design are significantly more effective than standard methods.

Is responsible for the preparation and teaching of engineering-oriented statistical analysis courses given by the Department.

Provides assistance as required in the development of data bases and companion software packages to extract statistical information from them.

Writes technical papers for formal presentation and publication describing unique problem-solving approaches developed by the Statistical Engineering Group.

Exercises initiative and judgment to suggest and develop new statistical approaches to perform various engineering and other functions more effectively.

Represents the Department in statistical matters at national, regional, and state conferences.

Maintains contact with university faculty members and other professionals to keep abreast of the latest advances in engineering applications of statistical analysis.

Develops and maintains a library of statistical literature to support group activities.

Supervises work operations and/or functional programs and has responsibility for effectively recommending the hiring, firing, promoting, demoting, and/or disciplining of employees.

REQUIREMENTS

Education

Graduation from an accredited college with a Bachelor's degree in Civil Engineering. A Professional Engineer 's license issued or validated by the New Jersey Board of Professional Engineers and Land Surveyors may be substituted for the Bachelor's degree in Engineering.

Graduation from an accredited college with a Master's degree in Statistical Analysis or Applied Statistics with a minimum of six credit hours in computer programming.

A minimum of 24 credit hours in applied statistics plus six credit hours in computer programming at the graduate or undergraduate level at an accredited college may be substituted for the Master's degree.

Experience

Four years of profession experience in the various modes of transportation engineering or engineering research, at least three years of which shall have been in a supervisory capacity, and at least three years of which shall include statistical work involving design of experiments, collection and analysis of data, hypothesis testing, development of statistical acceptance procedures, computer programming, writing of technical reports, and oral presentations.

A Doctor's degree in Statistical Analysis or Applied Statistics may be substituted for two years of the required experience in a non-supervisory capacity.

License

Appointee will be required to possess a driver's license valid in New Jersey only if the operation of a vehicle, rather than employee mobility, is necessary to perform the essential duties of this position.

Knowledges and Abilities

Thorough theoretical and practical knowledge of a broad range of statistical topics including probability theory, frequency distributions, sampling, confidence interval estimation, hypothesis testing, regression analysis, design of experiments, analysis of variance, contingency tables, goodness-of-fit tests, and non-parametric methods.

Thorough knowledge and understanding of variables and attributes acceptance plans, with variability known or unknown, and their associated operating characteristic curves.

Thorough knowledge of basic assumptions upon which all commonly used statistical procedures are based.

Wide knowledge of advanced computer programming techniques including modular design with emphasis on efficiency and clarity.

Wide knowledge of the techniques of computer simulation.

Wide knowledge of engineering materials, methods, equipment, and procedures related to all aspects of public transportation.

Wide knowledge of standard quality control tests, procedures, and measuring devices including familiarity with laboratory equipment and testing machines.

Considerable knowledge of operations research techniques (linear programming) and computer programs for their implementation.

Considerable knowledge of engineering analysis and design procedures including probabilistic design and reliability analysis.

Considerable knowledge of and ability to apply effective management techniques.

Ability to bring a broad range of statistical, mathematical, and engineering analysis techniques to bear upon a wide variety of complex engineering problems which may include applications related to quality control, design, maintenance, safety, legal evidence and testimony, and other functions.

Ability to recognize and identify potentially serious problems which might require special effort or outside assistance.

Ability to write and validate efficient, well documented, conversational computer programs.

Ability to apply the techniques of computer simulation to a wide variety of engineering and other problems.

Ability to review and evaluate statistical and engineering publications for possible applications to Departmental problems.

Ability to make effective oral presentations in order to conduct seminars, teach statistics courses, provide expert testimony, and when necessary, explain complex statistical techniques in layman’s terms.

Ability to plan and assign work for subordinate employees and to supervise and instruct them in the performance of their work.

Ability to stimulate and guide the creative energies of subordinates.

Ability to work harmoniously and in cooperation with all units of the Department of Transportation.

Ability to function independently with a minimum of supervision.

Ability to read, write, speak, understand, and communicate in English sufficiently to perform the essential functions of the job after reasonable accommodation is made for known limitations. If the accommodation cannot be made because it would cause the employer undue hardship, such persons may not be eligible.

NEW JERSEY DEPARTMENT OF CIVIL SERVICE

DIVISION OF CLASSIFICATION AND COMPENSATION

STATISTICAL ENGINEER II

DEFINITION

Under the direction of a Statistical Engineer I, or other supervisor, in the Division of Research and Demonstration, Department of Transportation, supervises and /or applies the more complex statistical and mathematical techniques to the solution of engineering and other problems; does related work as required.

EXAMPLES OF WORK

Under direction, independently carries out assignments in applying the more complex statistical and mathematical techniques such as hypothesis testing, determination of confidence limits, regression analysis, analysis of variance, goodness-of-fit tests, use of contingency tables, non-parametric tests, and others necessary to complete assigned tasks; gives professional guidance to subordinate employees.

Analyzes engineering tolerances required for various materials and construction specifications to assure that they are realistic from the standpoint of being obtainable, and effective from the standpoint of accomplishing the desired engineering function.

Prepares and writes statistically-oriented engineering specifications, establishing reasonable and balanced risks to promote smooth implementation and acceptance and reduce the likelihood of subsequent litigation.

Investigates various engineering applications of statistical analysis to assure that underlying theoretical assumptions are sufficiently satisfied.

Plans correlation and calibration studies of various engineering tests and measurement techniques and develops precision statements. Makes recommendations as to which methods are superior taking both statistical and engineering factors into account.

Under direction, writes the more complex special computer programs to develop and test the more advanced engineering applications of statistical analysis including the use of computer simulation and the writing of conversational programs; supervises the writing of the less complex computer programs.

Under direction, designs the more complex research experiments which include advising others concerning the types and quantity of data required and the methods of analysis to be used.

Prepares clear, well written, technically sound reports describing the engineering and statistical principles employed in various analyses.

Assigns, instructs, and supervises the work of others related to data collection and statistical computations.

Reviews, analyzes, and interprets engineering and statistical data and reports.

Keeps current with new developments and latest trends of thought in engineering applications of statistical analysis. Attends and assists in educational seminars and conferences.

Confers with staff members from other divisions of the Department to assess statistical needs and give guidance as to what techniques can most effectively be employed to solve various engineering and other problems.

Supervises work operations and/or functional programs and has responsibility for effectively recommending the hiring, firing, promoting, demoting, and/or disciplining of employees.

Under direction, prepares and teaches engineering-oriented statistical analysis courses given by the Department.

Maintains essential records and files of techniques and procedures used for various assignments.

REQUIREMENTS

Education

Graduation from an accredited college with a Bachelor's degree in Civil Engineering supplemented by a minimum of 18 credit hours in Applied Statistics at the graduate or undergraduate level at an accredited college plus 6 credit hours in Computer Programming.

A Professional Engineer's license issued or validated by the New Jersey Board of Professional Engineers and Land Surveyors may be substituted for the Bachelor's degree in Engineering.

Experience

Three years of professional experience in the various modes of transportation engineering or engineering research, or in engineering statistical work involving the collection and analysis of data, hypothesis testing, computer programming, and technical report writing.

A Master's degree in Statistical Analysis or Applied Statistics may be substituted for one year of the non-supervisory experience requirement.

A Doctor's degree in Statistical Analysis or Applied Statistics may be substituted for two years of the indicated experience in a non-supervisory capacity.

License

Appointee will be required to possess a driver's license valid in New Jersey only if the operation of a vehicle, rather than employee mobility, is necessary to perform the essential duties of this position.

Knowledges and Abilities

Wide knowledge of a broad range of statistical topics including probability theory, frequency distributions, sampling, confidence interval estimation, hypothesis testing, regression analysis, design of experiments, analysis of variance, contingency tables, goodness-of-fit tests, and non-parametric methods.

Wide knowledge of statistical principles to be able to develop original and creative ways to deal with a variety of engineering and other problems.

Wide knowledge of basic theoretical assumptions upon which the more commonly used statistical procedures are based.

Wide knowledge of computer programming including familiarity with computer simulation techniques.

Wide knowledge of engineering materials, methods, equipment, and procedures related to all aspects of public transportation.

Wide knowledge of standard quality control tests, procedures, and measuring devices including familiarity with laboratory equipment and testing machines.

Wide knowledge of statistical acceptance plans.

Basic knowledge of engineering analysis and design procedures including probabilistic design and reliability analysis.

Basic knowledge of the techniques of operations research (linear programming) and computer programs for their implementation.

Ability to recognize the applicability of statistical, engineering, and scientific concepts to the solution of transportation problems.

Ability to distinguish between statistical significance and practical significance from an engineering standpoint.

Ability to communicate clearly and effectively, both in written reports and in oral presentations.

Ability to plan and organize work in a logical and efficient manner.

Ability to review and evaluate statistical and engineering publications for possible application to Department problems.

Ability to apply the techniques of engineering-economics to make cost-benefit studies.

Ability to present the results of studies in a clear and concise manner using graphical or tabular techniques as appropriate.

Ability to give appropriate assignments and instructions to subordinate employees and supervise the performance of their work.

Ability to read, write, speak, understand, and communicate in English sufficiently to perform the duties of this position. American Sign Language or braille may also be considered as acceptable forms of communication.

Persons with mental or physical disabilities are eligible as long as they can perform the essential functions of the job after reasonable accommodation is made for known limitations. If the accommodation cannot be made because it would cause the employer undue hardship, such persons may not be eligible.

NEW JERSEY DEPARTMENT OF CIVIL SERVICE

DIVISION OF CLASSIFICATION AND COMPENSATION

STATISTICAL ENGINEER III

DEFINITION

Under the direction of a Statistical Engineer II, or other supervisor, in the Division of Research and Demonstration, Department of Transportation, applies complex statistical and mathematical techniques to the solution of engineering and other problems; does related work as required.

EXAMPLES OF WORK

Under direction, independently applies the appropriate statistical and mathematical techniques such as hypothesis testing, determination of confidence limits, regression analysis, analysis of variance, goodness–of–fit tests, use of contingency tables, non-parametric tests, and others necessary to complete assigned tasks.

Under direction, analyzes engineering tolerances required for various materials and construction specifications to assure that they are realistic from a standpoint of being obtainable, and effective from the standpoint of accomplishing the desired engineering function.

Under direction, prepares and writes statistically–oriented engineering specifications, establishing reasonable and balanced risks to promote smooth implementation and acceptance and reduce the likelihood of subsequent litigation.

Under direction, investigates various engineering applications of statistical analysis to assure that underlying theoretical assumptions are sufficiently satisfied.

Under direction, plans correlation and calibration studies of various engineering tests and measurement techniques and develops precision statements. Makes recommendations as to which methods are superior taking both statistical and engineering factors into account.

Under direction, writes special computer programs to develop and test various engineering applications of statistical analysis including the use of computer simulation and the writing of conversational programs.

Under direction, designs research experiments which include advising others concerning the type and quantity of data required and the methods of analysis to be used.

Prepares clear, well-written, technically sound reports describing the engineering and statistical principles employed in various analyses.

Assigns, instructs, and supervises the work of others related to data collection and statistical computations.

Reviews, analyzes, and interprets engineering and statistical data and reports.

Keeps current with new developments and latest trends of thought in engineering applications of statistical analysis. Attends and assists in educational seminars and conferences.

Assists in the preparation and teaching of engineering-oriented statistical analysis courses given by the Department.

Maintains essential records and files of techniques and procedures used for various assignments.

REQUIREMENTS

Education

Graduation from an accredited college with a Bachelor=s degree in Civil Engineering.

A Professional Engineer=s license issued or validated by the New Jersey Board of Professional Engineers and Land Surveyors may be substituted for the Bachelor=s degree in engineering.

A minimum of 12 credit hours in applied statistics at the graduate or undergraduate level at an accredited college plus 6 credit hours in Computer Programming.

Experience

Two years of professional experience in the various modes of transportation engineering or engineering research, or in engineering statistical work involving the collection and analysis of data, hypothesis testing, computer programming, and technical report writing.

A Master=s degree in statistical analysis or applied statistics may be substituted for one year of the required experience.

A Doctor=s degree in statistical analysis or applied statistics may be substituted for two years of the indicated experience in a non-supervisory capacity.

License

Appointee will be required to possess a driver=s license valid in New Jersey only if the operation of a vehicle, rather than employee mobility, is necessary to perform the essential duties of the position.

Knowledges and Abilities

Wide knowledge of a broad range of statistical topics including probability theory, frequency distributions, sampling, confidence interval estimation, hypothesis testing, regression analysis, design of experiments, analysis of variance, contingency tables, goodness-of-fit tests, and non-parametric methods.

Considerable knowledge of statistical principles to be able to develop original and creative ways to deal with a variety of engineering and other problems.

Considerable knowledge of basic theoretical assumptions upon which the more commonly used statistical procedures are based.

Considerable knowledge of computer programming.

Basic knowledge of engineering analysis and design procedures including probabilistic design and reliability analysis.

Basic knowledge of standard quality control tests, procedures, and measuring devices including familiarity with laboratory equipment and testing machines.

Basic knowledge of engineering materials, methods, equipment, and procedures related to public transportation.

Basic knowledge of statistical acceptance plans.

Ability to recognize the applicability of statistical, engineering, and scientific data to the solution of transportation problems.

Ability to distinguish between statistical significance and practical significance from an engineering standpoint.

Ability to communicate clearly and effectively, both in written reports and in oral presentations.

Ability to plan and organize work in a logical and efficient manner.

Ability to review and evaluate statistical and engineering publications for possible application to Department problems.

Ability to apply the techniques of engineering-economics to make cost-benefit studies.

Ability to present the results of studies in a clear and concise manner using graphical or tabular techniques as appropriate.

Ability to maintain essential records and files of techniques and procedures used fo various assignments.

Ability to read, write, speak, understand, or communicate in English sufficiently to perform the duties of this position. American Sign Language or Braille may also be considered as acceptable forms of communication.

Persons with mental or physical disabilities are eligible as long as they can perform the essential functions of the job after reasonable accommodation is made to their known limitations. If the accommodation cannot be made because it would cause the employer undue hardship, such persons may not be eligible.

REFERENCES

1. *Special Report 61B: AASHO Road Test, Report 2—Materials and Construction*, Highway Research Board, Publication 951, National Academy of Sciences-National Research Council, Washington, DC, 1962.

2. "Glossary of Highway Quality Assurance Terms," *Transportation Research Circular No. E-C037,* Transportation Research Board, Washington, DC, April 2002.

3. "Statistically Oriented End-Result Specifications," *NCHRP Synthesis 38*, Transportation Research Board, Washington, DC, 1976.

4. "Quality Assurance," *NCHRP Synthesis 65,* Transportation Research Board, Washington, DC, 1979.

5. "Performance-Related Specifications for Highway Construction and Rehabilitation," *NCHRP Synthesis 212,* Transportation Research Board, Washington, DC, 1995.

6. "23 CFR Part 637B," *Quality Assurance Procedures for Construction*, Federal Highway Administration, Federal Register, July 1995.

7. *AASHTO Quality Assurance Guide Specification*, AASHTO Quality Construction Task Force, American Association of State Highway and Transportation Officials, February 1996.

8. *Manual on Presentation of Data and Control Chart Analysis, Sixth Ed.*, ASTM Committee E-11 on Quality and Statistics, American Society for Testing and Materials, Philadelphia, PA, 1992.

9. *The Memory Jogger II*™, ISBN 1–879364–44–1, GOAL/QPC, Salem, NH, January 1994.

10. Grant, E.L. and Leavenworth, R.S., *Statistical Quality Control, Seventh Ed.*, ISBN 0–07–043555–3, McGraw–Hill, January 1996.

11. Burati, J.L. and Hughes, C.S., "Highway Materials Engineering Module I: Materials Control and Acceptance—Quality Assurance," *National Highway Institute Course No. 13123*, Federal Highway Administration, Washington, DC, 2001.

12. "AASHTO Draft Standard Recommended Practice for Technician Training and Qualification Programs," *AASHTO Materials, Part I Specifications: R–25–00*, American Association of State Highway and Transportation Officials, 2000.

13. "Program Detail Manual for Certification in the Field of Transportation Engineering Technology," *Subfield of Highway Materials, Fifth Ed.*, National Institute of Certification in Engineering Technologies, 1996.

14. "Program Detail Manual for Certification in the Field of Construction Materials Testing," *Subfield of Asphalt, Concrete, and Soils, Sixth Ed.*, National Institute of Certification in Engineering Technologies, 1997.

15. *Guidelines for Establishing a Technician Training and Certification Program,* National Quality Initiative, December 1997.

16. *AASHTO Implementation Manual for Quality Assurance*, AASHTO Quality Construction Task Force, American Association of State Highway and Transportation Officials, February 1996.

17. Burati, J.L., et al, unpublished findings.

18. Weed, R.M., "Quality Assurance Software for the Personal Computer," *Publication No. FHWA–SA–96–026*, Federal Highway Administration, Washington, DC, May 1996.

19. "The Use of Incentive and Disincentive Provisions Regarding Quality and Completion Time for Federal Aid Highway Construction," *National Experimental Evaluation Program Project 24, Notice N 5080.66*, Federal Highway Administration, Washington, DC, 1977.

20. Weed, R.M., "Unbiased Graduated Pay Schedules," *Transportation Research Record 745*, Transportation Research Board, Washington, DC, 1980.

21. Schmitt, R.L., Russell, J.S., Hanna, A.S., Bahia, H.U., and Jung, G.S., "Summary of Current Quality Control/Quality Assurance Practices for Hot-Mix Asphalt Concrete," *Transportation Research Record 1632*, Transportation Research Board, Washington, DC, 1998.

22. "Acceptance Sampling Plans for Highway Construction," *AASHTO Materials, Part I Specifications: R–9*, American Association of State Highway and Transportation Officials, 1984.

23. "Military Standard 414, Sampling Procedures for Inspection by Variables for Percent Defective," U.S. Department of Defense, Washington, DC, June 1957.

24. Weed, R.M., "Development of Air Voids Specification for Bituminous Concrete," *Transportation Research Record 1491*, Transportation Research Board, Washington, DC, 1995.

25. Palise, F., Strizki, B., and Weed, R.M., "Upgrading the Asphaltic Concrete Specification of the New Jersey Department of Transportation," *Transportation Research Record 1616*, Transportation Research Board, Washington, DC, 1998.

26. Sayers, M.W. and Karamihas, S.M., *The Little Book of Profiling*, The University of Michigan Transportation Research Institute, National Highway Institute, Federal Highway Administration, Washington, DC, September 1998.

27. Hughes, C.S., Simpson, A.L., Cominsky, R., Pendleton, O.J., Weed, R.M., and Wilson, T., "Measurement and Specification of Construction Quality, Volume I," *Final Report*, Federal Highway Administration, FHWA–RD–98–077, May 1998.

28. Pearson, E.S., and Hartley, H.O., eds., *Biometrika Tables for Statisticians, Volume I*, Cambridge University Press, New York, 1929.

29. Jones, T.A., *Skewness and Kurtosis as Criteria of Normality in Observed Frequency Distributions*, Esso Production Research, Report 69–5, 1969.

30. Burington, R.S., *Handbook of Mathematical Tables and Formulas*, McGraw–Hill, New York, 1973.

31. *AASHTO Guide for Design of Pavement Structures*, American Association of State Highway and Transportation Officials, Washington, DC, 1993.

32. Weed, R.M., "A Rational Method for Relating As-Built Quality to Pavement Performance and Value," *Transportation Research Record 1632*, Transportation Research Board, Washington, DC, 1998.

33. Linden, R.N., Mahoney, J.P., and Jackson, N.C., "Effect of Compaction on Asphalt Concrete Pavement," *Transportation Research Record 1217*, Transportation Research Board, Washington, DC, 1989.