ALTERNATIVES TO THE USE OF CONTRACTOR'S QUALITY CONTROL DATA FOR ACCEPTANCE AND PAYMENT PURPOSES

A Thesis

by

SUJAY SUDHIR WANI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

May 2010

Major Subject: Civil Engineering

ALTERNATIVES TO THE USE OF CONTRACTOR'S QUALITY CONTROL DATA FOR ACCEPTANCE AND PAYMENT PURPOSES

A Thesis

by

SUJAY SUDHIR WANI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Nasir Gharaibeh
Committee Members,	Roger Smith
	Webster West
Head of Department,	John Niedzwecki

May 2010

Major Subject: Civil Engineering

ABSTRACT

Alternatives to the Use of Contractor's Quality Control Data for Acceptance and Payment Purposes. (May 2010)

Sujay Sudhir Wani, B.E, Mumbai University

Chair of Advisory Committee: Dr. Nasir Gharaibeh

Currently, several state Departments of Transportation (DOTs) are using contractor test results, in conjunction with verification test results, for construction and materials acceptance purposes. While the reasons for using contractor test results for construction and materials acceptance purposes are real (essentially shortage of state DOT staff and intensive construction schedules), the practice itself has fundamental pitfalls. This research reveals the conceptual and technical pitfalls of using contractor test results for acceptance and payment purposes; identifies and ranks potential alternatives and improvements to the use of contractor test results for acceptance and payment purposes; and investigates the potential application of skip-lot sampling as a means for reducing acceptance sampling and testing for highway agencies.

DEDICATION

This thesis is dedicated to my family and all my friends.

ACKNOWLEDGEMENTS

First of all, I would like to thank my committee chair, Dr. Gharaibeh for his guidance and support throughout the course of this research. I would also like to thank my committee members, Dr. Smith and Dr. West, for their timely guidance and support.

Thanks also go to all my friends here at Texas A&M for making my stay in this country very pleasant. I would also like to thank my colleagues, the department faculty and staff for making my masters journey at Texas A&M University a great experience.

Finally, very special thanks to my parents for their continuous encouragement, support and love.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES x

1. INTRODUCTION 1

    1.1 Background 1
    1.2. Problem Statement 2
    1.3. Research Objective and Scope 3
    1.4. Report Organization 5

2. LITERATURE REVIEW 6

    2.1. Verification Process 6
        2.1.1. Samples and Verification 7
        2.1.2. Hypothesis Testing and Level of Significance 9
        2.1.3. F-test and T-test 9
        2.1.4. Errors 14
        2.1.5. Percent within Limits (PWL) and Pay Factor 15
    2.2. Summary of Relevant Previous Studies 20

3. STATISTICAL AND CONCEPTUAL PITFALLS OF USING CONTRACTOR'S TEST RESULTS FOR ACCEPTANCE PURPOSES 23

    3.1. Statistical Pitfalls: Unreliable Statistical Tests 23
        3.1.1. Case 1: Asphalt Content with Poor Mean Value 25
        3.1.2. Case 2: PCCP Slab Thickness with Poor Mean Value 32
    3.2. Probability Profile 35
    3.3. Effect of Contractor's Sample Size 39
    3.4. Conceptual Pitfalls: Intermingling Process Control and Product Acceptance 41

Page

4. ALTERNATIVES AND IMPROVEMENTS TO CONTRACTOR ACCEPTANCE TESTING 45

4.1. Workshop Overview 45
4.2. Alternatives and Improvements to Contractor Acceptance Testing 45
4.3. Evaluation Method of the Eighteen Potential Alternatives/Improvements 52
4.4. Results of the Evaluation 54

5. SKIP-LOT SAMPLING PLANS 59

5.1. Rationale and Background of Skip-lot Sampling Plan 59
5.2. Skip-lot Sampling Plan-1 (SkSP-1) 61
5.3. Skip-lot Sampling Plan-2 (SkSP-2) 63
5.4. An Example Application of SkSP-2 67

6. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS 75

6.1. Summary 75
6.2. Conclusions 76
6.3. Recommendations 77

REFERENCES 79

APPENDIX A 81

APPENDIX B 94

APPENDIX C 98

VITA 112

LIST OF FIGURES

		Page
Figure 1.1	Research Scope	4
Figure 2.1	Graphical Illustration of the Contractor Test Results Validation Process	8
Figure 2.2	Example of OC Curve	15
Figure 2.3	Expected Pay Factor Trend Figure	17
Figure 3.1	Process Flowchart	24
Figure 3.2	Probability of Detecting Reduced Standard Deviation of Asphalt Content	37
Figure 3.3	Probability of Detecting Increase in Mean of Asphalt Content	38
Figure 3.4	Probability of Detecting Reduced Standard Deviation of PCC Thickness	38
Figure 3.5	Probability of Detecting Increase in Mean of PCC Thickness	39
Figure 3.6	Product-Focused Model for Construction and Materials Acceptance	42
Figure 3.7	Process-Focused Model for Construction and Materials Quality Control	43
Figure 3.8	Expected Outcome of Product Acceptance	44
Figure 3.9	Expected Outcome of Process Control	44
Figure 5.1	Curves for Determining Values of AOQL for Given f and i, and vice versa.	63
Figure 5.2	A Sketch of a SkSP-2 Plan for Highway Construction and Materials Lots.	64
Figure 5.3	Operating Characteristics Curve for the Original Acceptance Plan	68

Page

Figure 5.4 Average Sampling Numbers (i.e. Sample Size) for Different Historical Quality Levels 73

Figure 5.5 Fraction of Lots Subject to Acceptance Inspection for Different Historical Quality Levels 74

Figure 5.6 Probability of Acceptance (Reference Plan) vs Probability of Acceptance (SkSP-2) 74

Figure B-1 Panel Member Responses to Workload Reducing Alternatives 94

Figure B-2 Panel Member Responses to Alternatives Suggested as Improvement to CAT 95

Figure B-3 Score of Each Alternative Suggested for Workload Reduction 96

Figure B-4 Score of Each Alternative Suggested as Improvement to CAT 97

LIST OF TABLES

		Page
Table 2.1	Quality Index for Estimating PWL for n = 1 to 10	18
Table 2.2	Quality Index for Estimating PWL for n > 10	19
Table 3.1	Authentic Test Results for Asphalt Content	26
Table 3.2	Manipulated QC Test Results for Asphalt Content by Reducing Standard Deviation	27
Table 3.3	Original and Manipulated Asphalt Content Data (by Reducing Standard Deviation)	29
Table 3.4	Manipulated QC Test Results for Asphalt Content by Increasing the Mean	30
Table 3.5	Original and Manipulated Asphalt Content Data (by Increasing the Mean Value)	31
Table 3.6	Authentic Test Results for PCC Slab Thickness	33
Table 3.7	Original and Manipulated PCCP Slab Thickness Data (by Decreasing SD)	34
Table 3.8	Original and Manipulated PCCP Slab Thickness Data (by Increasing the Mean Value)	35
Table 3.9	Probability of Detecting Difference in Standard Deviation.	36
Table 3.10	Maximum Undetected Decrease in Standard Deviation in Asphalt Content	40
Table 3.11	Maximum Undetected Increase in Mean in Asphalt Content	40
Table 3.12	Maximum Undetected Decrease in Standard Deviation in PCC Thickness	41
Table 3.13	Maximum Undetected Increase in Mean in PCC Thickness	41
Table 4.1	Initial Set of Alternatives and Improvements to Contractor Acceptance Testing	46

		Page
Table 4.2	Evaluation Criteria of Identified Alternatives	52
Table 4.3	Top 5 Alternatives Based on Overall Average Score	55
Table 4.4	Top 5 Alternatives Based on Criterion #1 (Potential for Reducing Agency's Workload)	56
Table 4.5	Top 5 Alternatives Based on Criterion #2 (Potential for Increasing Agency's Risk of Accepting Poor Quality Products)	57
Table 4.6	Top 5 Alternatives Based on Criterion #3 (Ease of Implementation)	57
Table 4.7	Top 5 Alternatives Based on the Worthiness of Further Study	58
Table 5.1	Summary of SkSP-2 Calculations for the Example Problem	73
Table A-1	Contractor's Asphalt Content Data for Sample Size Ration of 1:1 (Reducing SD)	81
Table A-2	Contractor's Asphalt Content Data for Sample Size Ration of 1:1 (Increasing Mean)	81
Table A-3	Contractor's Asphalt Content Data for Sample Size Ration of 1:2 (Reducing SD)	82
Table A-4	Contractor's Asphalt Content Data for Sample Size Ration of 1:2 (Increasing Mean)	83
Table A-5	Contractor's Asphalt Content Data for Sample Size Ration of 1:4 (Reducing SD)	84
Table A-6	Contractor's Asphalt Content Data for Sample Size Ration of 1:4 (Increasing Mean)	85
Table A-7	Contractor's Asphalt Content Data for Sample Size Ration of 1:10 (Reducing SD)	86
Table A-8	Contractor's Asphalt Content Data for Sample Size Ration of 1:10 (Increasing Mean)	87

Page

Table A-10 Contractor's PCC Thickness Data for Sample Size Ration of 1:1 (Reducing SD) 88

Table A-10 Contractor's PCC Thickness Data for Sample Size Ration of 1:1 (Increasing Mean) 88

Table A-11 Contractor's PCC Thickness Data for Sample Size Ration of 1:2 (Reducing SD) 89

Table A-12 Contractor's PCC Thickness Data for Sample Size Ration of 1:2 (Increasing Mean) 89

Table A-13 Contractor's PCC Thickness Data for Sample Size Ration of 1:4 (Reducing SD) 90

Table A-14 Contractor's PCC Thickness Data for Sample Size Ration of 1:4 (Increasing Mean) 91

Table A-15 Contractor's PCC Thickness Data for Sample Size Ration of 1:10 (Reducing SD) 92

Table A-16 Contractor's PCC Thickness Data for Sample Size Ration of 1:10 (Increasing Mean) 93

Table C-1 Constants for SkSP-2 Plans Having Acceptance Constant of 0 98

Table C-2 Constants for SkSP-2 Plans Having Acceptance Constant of 1 99

Table C-3 Constants for SkSP-2 Plans Having Acceptance Constant of 2 100

Table C-4 Constants for SkSP-2 Plans Having Acceptance Constant of 3 101

Table C-5 Constants for SkSP-2 Plans Having Acceptance Constant of 4 103

Table C-6 Constants for SkSP-2 Plans Having Acceptance Constant of 5 104

Table C-7 Constants for SkSP-2 Plans Having Acceptance Constant of 6 105

Table C-8 Constants for SkSP-2 Plans Having Acceptance Constant of 7 106

Table C-9 Constants for SkSP-2 Plans Having Acceptance Constant of 8 107

Page

Table C-10 Constants for SkSP-2 Plans Having Acceptance Constant of 9 109

Table C-11 Constants for SkSP-2 Plans Having Acceptance Constant of 10 110

1. INTRODUCTION

1.1. Background

Currently several state Departments of Transportation (DOTs) are using contractor test results, in conjunction with verification test results, for construction and materials acceptance purposes. The use of contractor test results in acceptance decisions, was codified in 1995 in Title 23, Part 637, Code of Federal Regulations (23 CFR Part 637), the Federal Highway Administration's (FHWA's) Quality Assurance Procedures for Construction *(1)*. The American Association of State Highway and Transportation Officials (AASHTO) has adopted this shift towards making the contractors responsible for quality control (QC) and to allow contractor-performed tests to be used in acceptance decisions *(2)*. This is documented in AASHTO's "Implementation Manual for Quality Assurance" and "Quality Assurance Guide Specifications" *(3)*.

In the 1980s and early 1990s, there was growing perception that a duplication of testing was taking place: QC testing performed by the contractor and acceptance testing performed by the agency. This perception, coupled with the emphasis on reducing the number of government personnel, have resulted in the development and adoption of 23 CFR Part 637 *(4)*. The DOTs stopped performing quality control tests since these policy

---

The thesis follows style of *Transportation Research Record*.

changes were adopted. Today, more sampling and testing responsibilities are shifted to contractors due to shortage of DOT personnel and intensive construction schedules (e.g., night and weekend construction).

1.2. Problem Statement

Making use of contractor's test results (ideally used for process control) in acceptance and payment decisions have somewhat helped state DOTs to deal with shrinking workforce and lessen the workload involved in acceptance sampling and testing. But the practice itself has been controversial. Research has shown that the issue of bias in contractor test results is a concern *(2)*. Data gathered as part of the National Cooperative Highway Research Program (NCHRP) study 10-58(02) have shown the need for improvements and alternatives to this practice. NCHRP study 10-58(02) showed that there was a pattern of favorable contractor test results (in terms of both variability and mean values) for hot-mix asphalt concrete (HMAC). For Portland cement concrete (PCC) pavement and granular base, the results of the NCHRP study were less conclusive due to limited data. A subsequent study by LaVassar et al. *(5)* has questioned the findings of NCHRP 10-58(02) on the basis that grouping the data at the state DOT level leads to overly large sample size; and thus the statistical tests become much too discriminating to be used. Regardless of the data and analysis methods used in assessing the use of contractor test results for acceptance purposes, this practice has conceptual and technical pitfalls that need to be addressed. These pitfalls include:

- By allowing the intermingling of acceptance tests and quality control tests, this practice encourages a quality control approach that focuses on defect detection and containment rather than defect prevention.
- Creates an environment in which fraud is difficult to detect. This is because the statistical methods used for authenticating the contractor acceptance test results are unreliable for most practical testing frequencies.

Accordingly, there is a need to identify alternative testing and inspection strategies which can be used to reduce the acceptance sampling and testing workload without compromising the rigor of the quality assurance process.

1.3. Research Objective and Scope

The primary objectives of the research are as follows:

- Reveal the conceptual and technical pitfalls of using contractor test results for acceptance and payment purposes.
- Identify potential alternatives and improvements to the use of contractor test results for acceptance and payment purposes.
- Rank the identified alternatives and improvements to highlight the most promising strategies.
- Investigate the potential application of skip-lot sampling as a means for reducing acceptance sampling and testing.

To achieve the above objectives, a number of research activities were carried out as shown in Figure 1.1:

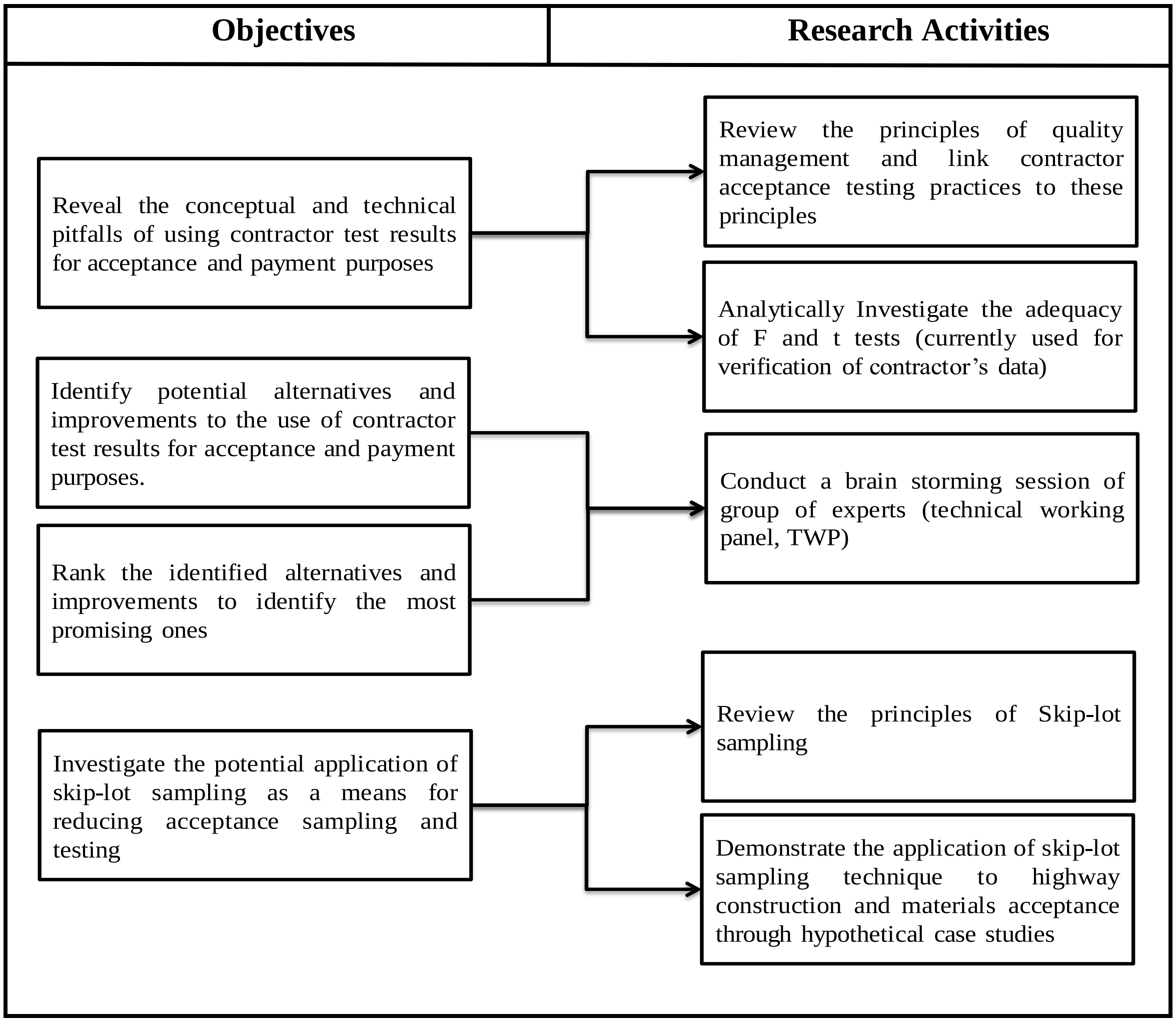

Figure 1.1: Research Scope

1.4. Report Organization

This research report is divided into six sections.

- Section 1 focuses on the background of the research problem and describes the research objectives and scope.
- Section 2 presents the literature review, focusing on current verification processes (specifically F and t tests) along with the concepts of percent-within limits (PWL) and pay adjustment.
- Section 3 sheds light on how statistical tests used for verification fail to detect potential manipulation of test results.
- Section 4 discusses and ranks potential alternatives and improvements to the use of contractor test results for acceptance purposes.
- Section 5 presents the method of skip-lot sampling and discusses its potential use as a technique for reducing sampling and testing workload for highway agencies.
- Section 6 presents conclusion and recommendation.

2. LITERATURE REVIEW

This section provides an overview of key concepts in acceptance sampling plans, statistical verification methods (t-test and F-test), and relevant studies on the use of contractor testing results for acceptance purposes.

2.1 Verification Process

When the contractor test results are used for product acceptance purposes, the agency (i.e., the buyer) should confirm or refute the acceptance test results using a reliable procedure. In construction projects that are partially funded through the federal government, Ruling 23 CFR 637 requires that verification testing be done by the agency *(6)*. 23 CFR 637 allowed contractor test results to be used in the acceptance decision provided that *(1)*:

- The sampling and testing has been performed by qualified laboratories and qualified sampling and testing personnel.
- The quality of the material has been validated by verification sampling and testing. The verification testing shall be performed on samples that are taken independently of the quality control samples.
- The quality control sampling and testing is evaluated by an independent assurance (IA) program.

2.1.1 Samples and Verification

According to FHWA, all the samples that are used for the process of quality control and acceptance decision should be chosen randomly. Each state DOT then decides on whether to use split samples or independent samples for verification purposes depending on the source of variability.

The agency is required or advised to carry out verification tests in order to minimize the risk of using biased contractor's data in acceptance decision. Some of the statistical measures adopted by agencies for verification purposes are as follow *(6)*:

- Verification of test strip testing at the beginning of the project,
- Validation of tests carried out by contractor with sporadic scrutinizing,
- Using statistical techniques to determine the variability of the test results, and
- Laying out a system for disputing the results with the consequences to be faced for noncompliance with acceptance data

AASHTO's Implementation manual of Quality Assurance helps to reduce the confusion regarding the split and independent samples, where the terms 'validation' and 'verification' are used to distinguish between 'independent samples' and 'split sample' respectively (5). Independent sample can be combined for acceptance decision if found statistically similar while split samples cannot be combined under any situation.

Generally, the procedures used by state DOTs to verify and validate contractor acceptance test results vary greatly from agency to agency *(4, 2)*. Currently, two verification procedures that are normally used for split samples are Paired t-test and D2S limit methods. D2S limit methods is very simple method which can be applied only in two test samples while with paired t-test number of pairs of split samples can be verified. Statistically robust t-test and F-test (which will be explained in greater detail in subsequent sections of this thesis) are used to validate the independent samples (Killingsworth and Hughes, 2002). Also, these statistical tests are recommended in the AASHTO *Implementation Manual for Quality Assurance (3)*. The AASHTO (1996) F-test and t-test validation process is illustrated in Figure 2.1.

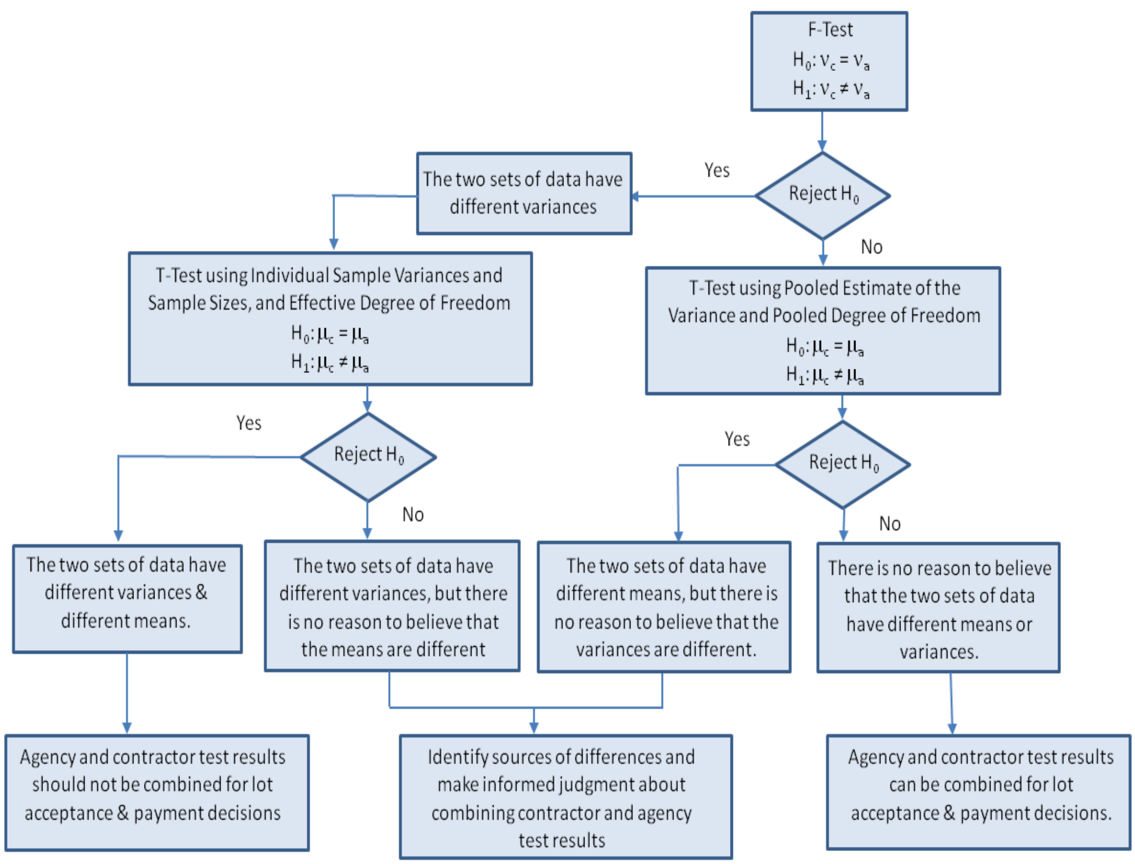

Figure 2.1: Graphical Illustration of the Contractor Test Results Validation Process *(3)*.

2.1.2 Hypothesis Testing and Level of Significance

It is important to understand the concept of hypothesis testing and level of significance before moving ahead to the discussion of various statistical procedures that are used for process or test method verification. Hypothesis tests are carried out whenever it is necessary to make a decision to accept or reject an assumption (e.g., mean or standard deviation of the two data sets is equal) made about two sets of data. Hypothesis testing checks the sample data against a claim or an assumption about the population. The claim or assumption made is termed as null hypothesis, H_0. Alternative claim or alternative hypothesis is another set of condition that is believed to be true, if the null hypothesis is rejected. A hypothesis test neither proves nor disapproves any assumption; it merely presents a formal statistical way by which any decision regarding the correctness of the assumption can be made *(7)*.

A select level of significance is assumed whenever a hypothesis test is carried out. This level of significance, α, is the probability of incorrectly rejecting the null hypothesis when it is true *(7)*. Typical values of α, are 0.10, 0.05, and 0.01. For example, if a null hypothesis is rejected at level of significance $\alpha = 0.05$, it signifies that there is only 5% chance that the hypothesis would be rejected in error when it is actually true.

2.1.3 F-test and T-test

When comparing the contractor's data and agency's data, the null hypothesis is that the data sets have come from the same population. In other words, the null hypothesis is that there is no difference between the variability of the data and there is no difference between the mean values of the data. The F-test provides a method by which the variances of the two data sets can be compared, while a t-test provides a method by which the mean of the two data sets can be compared. The construction processes and material properties usually follow a normal distribution *(7)*. The ratio of variances follow an F-distribution and the means of smaller sample sizes follow a t-distribution. Tables for both F-distributions and t-distributions are available in most statistics textbooks. Level of significance is chosen as 0.10, 0.05 or 0.01 before commencement of the tests. Though it is possible to carry out F-tests and t-tests manually with the help of tables and formula, it is advisable to use computer programs which simplifies the task and saves calculation time *(7)*.

2.1.3.1 F-test for Variance

In the context of verification testing, the basic purpose of conducting F-test is to check if there is a significant difference between variability in the contractor's test results and variability in the agency's test results. The F-test is generally carried out before the t-test as the outcome of this test has a significant impact on the way t-tests would be carried out (as discussed in the following section of this thesis). The null hypothesis for F-test is that there is no difference between the variance of the two data sets being compared. After comparing the variability in two sets of data with the help of an F-test, one can come to either of the following conclusions:

- Reject the hypothesis that there is no difference between the variability of the two sets of data

- Fail to reject the hypothesis that there is no difference between the variability of two sets of data

The Steps involved in carrying out F-tests are as follow *(7)*:

- Calculate the variances of both sets of data (i.e. calculate the variance of data provided by contractor and by agency). Let the variance of contractor's data be termed as s_c^2 and let variance of agency's data be termed as s_a^2
- Calculate the F-statistic as $F = s_c^2 / s_a^2$ or $F = s_a^2 / s_c^2$. Keep the larger value at the numerator so that F would always be greater than 1.
- Select the level of significance as 0.01, 0.05 or 0.10
- Determine the value of $F_{critical}$ from F-table (available in many statistical books) using the chosen level of significance and the degrees of freedom (n-1) of each set of data. The table given is a two sided table which detects if the variability of the two sets of data is different.
- Compare F-statistic with $F_{critical}$. If $F \geq F_{critical}$, then reject the null hypothesis that the variability of the two sets of data is same. In other words, the two sets of data have statistically different variability. If $F \leq F_{critical}$, then we fail to reject the null hypothesis. In other words, there is no evidence that the variabilities in the two data sets are significantly different from each other.
- p-value is defined as "the probability, if the test statistic really were distributed as it would be under the null hypothesis, of observing a test statistic [as extreme as, or more extreme than] the one actually observed" *(8)*. It is an independent value which can also be used in hypothesis testing (Used for the examples shown in

following sections). Null hypothesis is rejected if the p-value is smaller than or equal to the level of significance.

2.1.3.2 t-test for Mean

In the context of verification testing, the basic purpose of carrying out a t-test is to check if there is a significant difference between the mean of the contractor's test results and the mean of the agency's test results. The null hypothesis for a t-test is that there is no difference between the mean of the two data sets being compared. An F-test is generally carried out before the t-test as the outcome of this test has a significant impact on the way t-tests would be carried out. Depending on the outcome of the F-test, the t-test is carried out considering either equal variances or unequal variances. If the variances are found to be equal, the t-test is conducted by calculating a pooled (combining the variances of both data sets) variance and a pooled degree of freedom. If the variances are not found to be equal then the t-test is completed by considering individual variances and individual degrees of freedom for each data set being evaluated. After comparing the mean of two sets of data with the help of the t-test, one can come to either of the following conclusions:

- Reject the hypothesis that there is no difference between the mean of two sets of data
- Fail to reject the hypothesis that there is no difference between the mean of two sets of data

The steps involved in carrying out a t-test are as follows *(7)*:

- Calculate the means of the two data sets. Let the mean of contractor's data be termed as X_c and let the mean of agency's data be termed as X_a. Calculate the variances of the data sets. Term them as s_c^2 and s_a^2. Carry out the F-test on the data sets to determine if the variances are equal. Let n_c be the contractor's sample size and n_a be the agency's sample size.
- Depending on the result of the F-test, decide whether to pool the variances or use them individually. The formula used for pooling the variance is as follows:

$$Sp2 = Sc2 \times (\mathrm{nc} - 1) + Sa2 \times (\mathrm{na} - 1) / (\mathrm{nc} + \mathrm{na} - 2) \quad \text{... Eq. 2.1}$$

- Compute t-statistics. Following are the formulas used to compute t-statistic.
  - For equal variances:

$$t = |Xc - Xa| / \{\sqrt{[(sp2/nc) + (sp2/na)]}\} \quad \text{... Eq. 2.2}$$

  - For unequal variances:

$$t = |Xc - Xa| / \{\sqrt{[(Sc2/nc) + (Sa2/na)]}\} \quad \text{... Eq. 2.3}$$

- Obtain t-critical value from standard tables, with effective degrees of freedom computed as:
  - For equal variances:

$$\text{Degree of freedom} = (nc + na - 2) \quad \text{... Eq. 2.4}$$

  - For unequal variances:

$$\text{Degree of freedom} = \{[(Sc2/nc) + (Sa2/na)]2 / \{[(Sc2/nc)2/(nc + 1)] + [(Sa2/na)2/(na + 1)]\}\} - 2 \quad \text{... Eq. 2.5}$$

- Compare t-statistic with $t_{critical}$. If $t \geq t_{critical}$, then reject the null hypothesis that the mean of the two sets of data is same. In other words, the two sets of data have statistically different mean values. If $t \leq t_{critical}$, then the null hypothesis cannot be rejected. In other words, there is no evidence that the mean of the two data sets is significantly different from each other.

2.1.4 Errors

The chance or probability of making a correct decision when comparing two sets of data increases with increase in number of test results. In certain instances the F and t tests fail to detect a difference when it exists; while in certain instances F and t tests detect a difference between two data sets when it does not exist. Operating Characteristics (OC) curves help determine the total number of samples needed to achieve a particular probability of acceptance. OC curves plot the probability of detecting a difference (or probability of not detecting a difference) versus the actual difference in the studied statistic between the two data sets *(7)*.

The risk of incorrectly detecting a difference between two data sets (when in fact it does not exist) is known as Type I or α error (also called seller's risk). The risk of not detecting a differennce (when in fact it exists) is known as Type II or β error (also called buyer's risk). Figure 2.2 shows the OC curve of not detecting a difference when it actually exists. It can be observed that if the actual difference between the two data sets is 0 (i.e. there is no difference between the two data sets), there is 0.95 probability that the F and t tests would not detect a difference. But there is 5% chance that a difference would be detected in such cases. This 5% chance is known as Type I or α error. Also, consider the actual difference to be 0.5 units if the sample size is 10, there is 0.7

probability that this difference would not be detected by F and t tests. This 70% chance of not detecting difference is known as Type II or β error. It can be observed that the probability of not detecting a difference decreases with increase in sample size *(7)*.

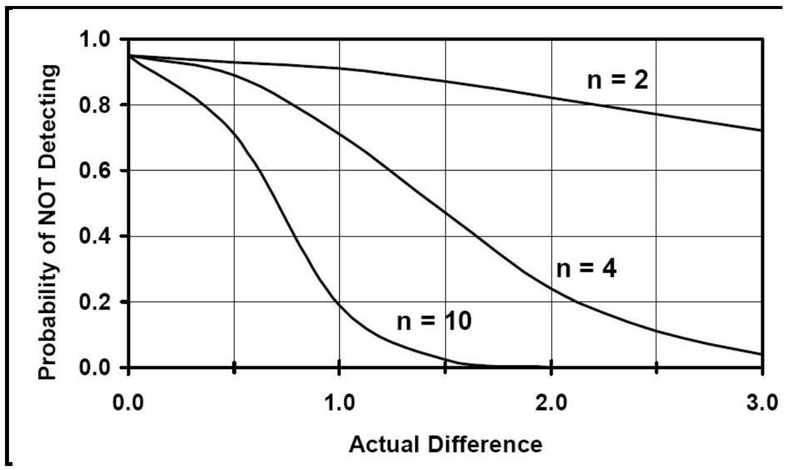

Figure 2.2: Example of OC Curve *(7)*

2.1.5 Percent within Limits (PWL) and Pay Factor

The overall quality of the delivered construction material can be measured using the percent within limits (PWL) statistic. The Transportation Research Board (TRB) glossary *(9)* defines PWL as "The percentage of the lot falling above the lower specification limit (LSL), beneath the upper specification limit (USL), or between the USL and LSL. [PWL may refer to either the population value or the sample estimate of the population value." Some DOTs use the statistic percent defective; where PD=100 – PWL. In this method the DOT decides the lower and the upper specification limits by using the mean and standard deviation. The limit is normally set at two standard deviations from the mean towards each side *(10)*. The percent of lots that is within the specifications set is then determined by this method.

The PWL procedure is conceptually based on normal distribution. The area under the curve is calculated to determine the percent of population lying within certain limits. Quality Index (instead of z-statistic) is the parameter that is used to determine the PWL. Quality Index (Q) is given by following formulas *(7)*

$$QL = \frac{(Xavg - LSL)}{s} \quad \text{... Eq. 2.6}$$

$$QU = \frac{(USL - Xavg)}{S} \quad \text{... Eq. 2.7}$$

where,

Q_L = quality index for lower specification limit

Q_U = quality index for upper specification limit

LSL = lower specification limit

USL = upper specification limit

s = standard deviation

X_{avg} = mean of sample.

Quality Index is used with along with PWL standard tables (see Tables 2.1 and 2.2) to calculate the PWL estimates. Q_L and Q_U are used to calculate the PWL for lower specification limit and upper specification limit, respectively. PWL for any particular set of data is then calculated by using following equation *(7)*:

$$PWL = PWL_L + PWL_U - 100 \quad \text{... Eq. 2.8}$$

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 = percent within the upper and lower specification limits.

Statistical acceptance plans determine pay factor (PF) as a function of PWL; where PF is a percentage of bid price. For example, AASHTO's pay factor formula is shown below:

$$\text{Pay Factor} = 55 + 0.5*\text{PWL} \quad \text{... Eq. 2.9}$$

As shown in Figure 2.3, using this pay factor curve, the contractor receives 100% pay for lots that have PWL of 90% (this PWL level is termed acceptable quality level or AQL). A lot is considered of poor quality if PWL is at (or below) the rejectable quality level (RQL). Typically, an AQL of 90% and RQL of 50% are used by DOTs.

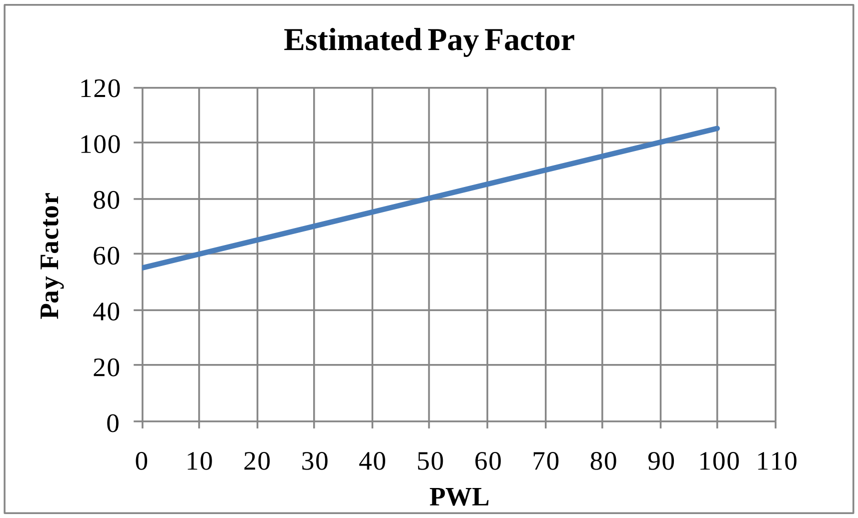

Figure 2.3: Expected Pay Factor Trend Figure

Table 2.1: Quality Index for Estimating PWL for n = 1to 10 *(7)*

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

Table 2.2: Quality Index for Estimating PWL for n > 10 ***(7)***

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

2.2 Summary of Relevant Previous Studies

Various other studies have been carried out previously to compare the contractor's data and the agency's data and assess the differences between the two. Some of these studies and their results are discussed herein.

Hancher et al. *(11)* discussed the possible advantages of and concerns about using contractor data for acceptance purposes. Possible advantages include coping with the reduction in state personnel, making contractor responsible for their own product, and improving dispute resolution. The concerns include the validity of the test data, contractor operating at lower end of specifications and lack of understanding of the process *(11)*.

Parker and Turochy *(2)* carried out a study which focused on comparing HMA test results from contractors and agencies. Testing data from HMA projects for entire construction season was collected from Florida, North Carolina and Kansas state DOTs. The statistical comparisons of variability (i.e., standard deviation) and central tendency (i.e. mean) for independent samples were then carried out using F-test and t-test at a significance level of 1% while the split samples were compared using paired t-test. The analyses showed a similar trend across the three states: regardless of the sampling procedure (split or independent) or the test type (HMA density obtained from nuclear gauges or core), the contractor results across the states were closer to the target value and were less variable than agency results. In most cases, the difference in deviation was statistically significant. Because these consistent statistical differences were found, parker and Turochy suggested that the use of contractor data in acceptance procedures should be limited.

LaVassar et al. *(5)* examined the practice of incorporating contractor's QC results in acceptance procedures and determining pay factor, by mainly determining the percentage of state DOT projects in which there is a significant statistical difference between contractor's QC results and DOT's QA results. The researchers utilized statistical measures like F-tests and t-tests. Authors used data provided by California, Minnesota, Texas and Washington state DOTs. These analyses were carried out in accordance with specifications or the methods used by the individual agencies. In the end, authors concluded that if the results are analyzed at statewide level, the statistical tests are much too discriminating to be used. Also the average number of projects or parameters that depicted significant statistical differences between the contractor's and agency's data data was fairly consistent among the studied states *(5)*.

Killingsworth and Hughes *(6)* suggested that if contractor's data is to be used for acceptance and payment decisions, contractor prequalification procedures should be developed (such as the procedure followed by Ontario Ministry of Transportation). In the end, they concluded by saying that there can be two major impacts of using contractor's data: 1) psychological impact on agency personnel of trusting contractor's data and 2) the need of implementing and checking a sound validation system *(6)*.

As can be seen from the above discussion, there is a general agreement that there is a need to check the validation process and to identify some alternatives and improvements to the practice of using contractor's test data in acceptance and pay decisions. This research is a step forward in filling this gap in the literature.

3. STATISTICAL AND CONCEPTUAL PITFALLS OF USING CONTRACTOR'S TEST RESULTS FOR ACCEPTANCE PURPOSES

3.1. Statistical Pitfalls: Unreliable Statistical Tests

This section demonstrates the inability of the F-test and t-test to detect manipulation in test results. The cases used here are based on hypothetical, yet plausible, test results. It was assumed that the contractor is producing poor quality materials; thus the population mean is close to the lower specification limit. Cases of data manipulation were simulated, where the contractor manipulates test results to increase the pay factor. Finally, the F-test and t-test is performed on these simulated cases of manipulated contractor's data and original agency data. A 0.05 level of significance is used in this analysis. The process used for the analysis is explained in Figure 3.1. The analysis was carried out considering various combinations of sample sizes. Ratios of 1:1, 1:2, 1:4, 1:10 (agency to contractor) were considered. Two cases are presented hereby to demonstrate the inability of the statistical tests to detect the manipulation in the test results. The data generated for other cases is given in Appendix A.

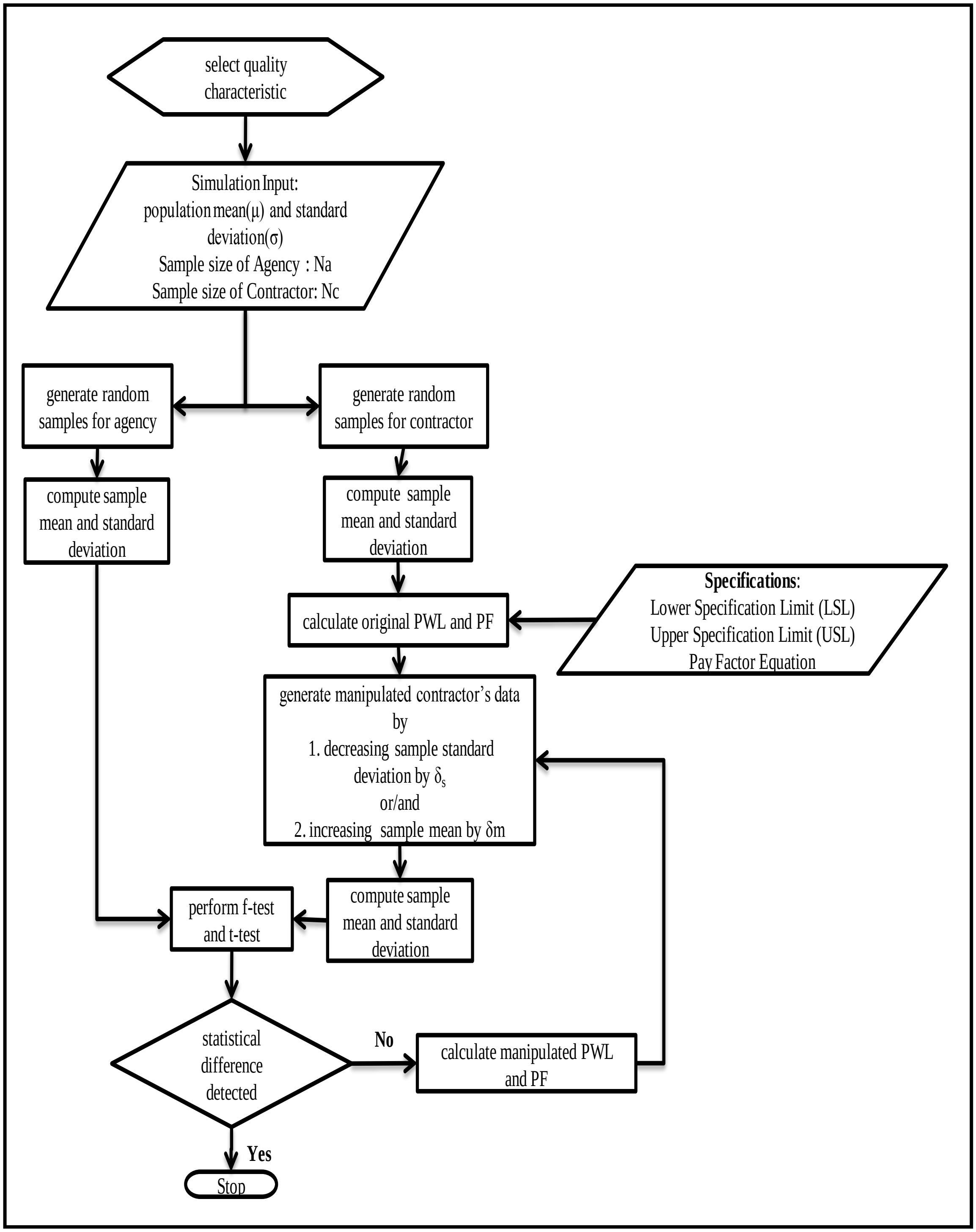

Figure 3.1: Process Flowchart

3.1.1. Case 1: Asphalt Content with Poor Mean Value

This case assumes that the acceptance quality characteristic (AQC) under consideration is asphalt content. The specifications parameters (reference sampling plan) are as follows:

- Target Mean Value: 5.22 percent
- Standard Deviation: 0.22 percent
- Lower specification limit (LSL): 4.82 percent
- Upper specification limit (USL): 5.62 percent
- Quality Measure: percent within limit (PWL)
- Pay Equation: PF = 55 + 0.5 PWL

Suppose that the contractor produced low asphalt content (i.e., low mean value) at a typical standard deviation. Thus, suppose that the true (as-built) asphalt content has the following mean and standard deviation:

- True asphalt content mean (μ): 4.9 percent
- True asphalt content standard deviation (σ): 0.22 percent

Suppose that the contractor's QC results consist of 10 tests and the independent validation results consist of five acceptance tests (all obtained from the above population). Computer simulation was used here to generate these test results randomly from the normal distribution of the above population. The original (i.e., authentic) test results are shown in Table 3.1.

Table 3.1: Authentic Test Results for Asphalt Content

QC Authentic Test results (Asphalt Content, %)	Independent Validation Test Results (Asphalt Content, %)
5.02	4.64
5.18	5.20
4.52	4.79
4.88	4.86
4.95	5.04
4.97	
5.22	
4.71	
4.80	
4.78	

The sample statistics, PWL, and pay factor (PF) for the above authentic test results (combined) are as follows:

- Sample Mean = 4.902%
- Sample Standard Deviation = 0.213%
- Lower Quality Index (QIL) = 0.400
- Upper Quality Index (QIU) = 3.479
- PWL = 65.3%
- Authentic PF = 87.65%

The above pay factor (87.65 percent) can be increased by manipulating the QC test results by either increasing the mean value or reducing the standard deviation. These two cases are discussed in the following sections.

- **Reduce the Standard Deviation**

Manipulated test results were generated randomly (through computer simulation) from a normal distribution with a mean value of 4.9 percent (i.e., unchanged from the authentic mean value) and a standard deviation of 0.18 percent (i.e., reduced from the authentic standard deviation of 0.22 percent). The manipulated set of data is shown in Table 3.2

Table 3.2: Manipulated QC Test Results for Asphalt Content by Reducing Standard Deviation

Manipulated Test results (Asphalt Content, %)
5.03
5.08
4.66
4.87
4.91
4.96
5.24
4.72
4.82
4.77

The sample statistics, PWL, and pay factor for the new combined data (manipulated QC test results and authentic validation test results) are as follows:

- Sample Mean = 4.906%
- Sample Standard Deviation = 0.177%
- Lower Quality Index (QIL) = 0.465

- Upper Quality Index (QIU) = 3.898
- PWL = 67.9%
- PF = 88.95

It can be seen that the pay factor has increased from 87.65 percent to 88.95 percent. The question is can the F-test detect this manipulation?

For the F-test, the p-value = 0.562. Since p-value > α (i.e., 0.562 > 0.05), there is no reason to conclude that the sample variances are not equal. Thus, at a significance level of 0.05 (α=0.05), the F-test was unable to detect the manipulation. Indeed, the QC results can be manipulated even further (i.e., the sample standard deviation can be reduced to 0.105 percent) without being detected by the F-test (see Table 3.3). When the standard deviation was reduced to 0.098 percent, however, the F-test results show that the two data sets have different variance (i.e., the F-test detected this data manipulation).

Table 3.3: Original and Manipulated Asphalt Content Data (by Reducing Standard Deviation)

Test No.	Original Test Results	Manipulated Test Results				
	SD=0.213%	Case 1 SD= 0.207%	Case 2 SD =0.177%	Case 3 SD=0.149%	Case 4 SD =0.105%	Case 5 SD= 0.098%
1	5.02	5.06	5.03	4.97	4.97	4.94
2	5.18	5.07	5.08	5.07	5.02	5.00
3	4.52	4.62	4.66	4.73	4.74	4.76
4	4.88	4.90	4.87	4.88	4.89	4.88
5	4.95	4.94	4.91	4.93	4.92	4.92
6	4.97	4.98	4.96	4.97	4.98	4.97
7	5.22	5.32	5.24	5.18	5.07	5.07
8	4.71	4.67	4.72	4.71	4.79	4.79
9	4.80	4.84	4.82	4.83	4.86	4.86
10	4.78	4.78	4.77	4.79	4.83	4.82
Sample Mean	4.902	4.918	4.906	4.906	4.907	4.901
Sample Std. Dev.	0.213	0.207	0.177	0.149	0.105	0.098
p-value for F-test	0.877	0.826	0.562	0.322	0.065	0.044
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**	**Yes**
p-value for t-test	0.985	0.905	0.985	0.984	0.971	0.970
Reject H0 for t-test?	**No**	**No**	**No**	**No**	**No**	**No**
Quality Index (L)	0.400	0.461	0.465	0.512	0.559	NA
Quality Index (U)	3.479	3.490	3.898	4.291	5.028	NA
PWL(L)	65.3	67.3	67.9	69.3	71	NA
PWL(U)	100	100	100	100	100	NA
PWL	65.3	67.3	67.9	69.3	71	NA
Pay Factor	**87.65**	**88.65**	**88.95**	**89.65**	**90.5**	**NA**

**SD: Standard Deviation, *NA: Not Applicable (difference detected)*

For the t-test also, the p-value = 0.985. Since p-value > α (i.e., 0.984 > 0.05), there is no reason to assume that the sample means are not equal. The t-test result is expected since the mean value was not manipulated.

- **Increase the Mean Value**

Manipulated QC test results were generated randomly (through computer simulation) from a normal distribution with a mean value of 5.1 percent (i.e., increased from the authentic mean value of 4.9 percent) and a standard deviation of 0.22 percent (i.e., unchanged from the authentic standard deviation). The manipulated set of data is shown in Table 3.4

Table 3.4: Manipulated QC Test Results for Asphalt Content by Increasing the Mean

Manipulated QC Test results (Asphalt Content, %)
4.99
5.25
4.96
4.90
5.21
5.05
5.33
5.39
4.67
5.11

The sample statistics, PWL, and pay factor for the new combined data (manipulated QC test results and authentic validation test results) are as follows:

- Sample Mean = 5.085%
- Sample Standard Deviation = 0.218%
- Lower Quality Index (QIL) = 0.896
- Upper Quality Index (QIU) = 2.605
- PWL = 81.0%
- PF = 95.5%

It can be seen that the pay factor has increased from 87.15 percent to 95.5 percent. The question is can the t-test detect this manipulation?

For the t-test, the p-value = 0.153. Since p-value > α (i.e., 0.153 > 0.05), there is no reason to conclude that the sample means are not equal. Thus, at a significance level of 0.05 (α=0.05), the t-test was unable to detect the data manipulation to increase the mean value of the QC results. Indeed, the QC results can be manipulated even further (i.e., the sample mean can be increased to 5.15 percent) without being detected by the t-test (see Table 3.5). When the mean was increased to 5.17 percent, however, the t-test detected this manipulation.

Table 3.5: Original and Manipulated Asphalt Content Data (by Increasing the Mean Value)

Test No.	Original Test Results	Manipulated Data				
	Mean=4.902%	Case 1 Mean = 5.085%	Case 2 Mean = 5.115%	Case 3 Mean = 5.130%	Case 4 Mean = 5.149%	Case 5 Mean= 5.169%
1	5.02	4.99	5.04	5.04	5.03	4.88
2	5.18	5.25	5.20	5.12	4.94	5.00
3	4.52	4.96	5.08	4.97	5.21	5.30
4	4.88	4.90	5.29	5.33	5.30	4.91
5	4.95	5.21	4.83	5.15	4.99	5.09
6	4.97	5.05	4.95	5.54	5.38	5.18
7	5.22	5.33	5.49	5.25	5.56	5.14
8	4.71	5.39	5.16	5.19	5.12	5.26
9	4.80	4.67	4.76	4.90	5.17	5.37
10	4.78	5.11	5.34	4.80	4.79	5.57
Sample Mean	4.902	5.085	5.115	5.130	5.149	5.169
Sample Std. Dev.	0.213	0.218	0.229	0.217	0.228	0.214
p-value for F-test	0.857	0.903	0.985	0.890	0.984	0.866
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**	**No**
p-value for t-test	0.988	0.153	0.111	0.080	0.068	0.042
Reject H_0 for t-test?	**No**	**No**	**No**	**No**	**No**	**Yes**
Quality Index (L)	0.420	0.896	0.932	0.989	0.996	NA
Quality Index (U)	3.496	2.605	2.391	2.389	2.227	NA
PWL(L)	65.3	81.5	82.25	83.75	84	NA
PWL(U)	100	99.5	99.3	99.3	99.1	NA
PWL	65.3	81	81.55	83.05	83.1	NA
Pay Factor	**87.65**	**95.5**	**95.775**	**96.525**	**96.55**	NA

**NA: Not Applicable (difference detected)*

3.1.2. Case 2: PCCP Slab Thickness with Poor Mean Value

This case assumes that the AQC under consideration is slab thickness of Portland cement concrete pavement (PCCP).

The specifications parameters are as follows:

- Target Mean Value: 12 inches
- Lower specification limit (LSL): 11 inches
- Quality Measure: percent within limit (PWL)
- Pay Equation: PF = 55 + 0.5 PWL

Suppose that the contractor produced low PCCP slab thickness (i.e., low mean value) at a typical standard deviation. Thus, suppose that the true (as-built) PCCP slab thickness has the following mean and standard deviation:

- True mean PCCP slab thickness (μ): 11.25 inches
- True standard deviation (σ): 0.4 inches

Suppose that the QC results consist of 8 tests and the independent verification results consist of four acceptance tests (all obtained from the above population). Computer simulation was used here to generate these test results randomly from the above normal distribution of the above population. The original (i.e., authentic) test results are shown in Table 3.6.

Table 3.6: Authentic Test Results for PCCP Slab Thickness

QC Authentic Test results (PCCP Slab Thickness, in)	Independent Validation Test Results (PCCP Slab Thickness, in)
11.27	11.77
10.68	10.86
11.49	10.99
10.81	11.41
11.07	
11.18	
11.91	
11.58	

The sample statistics, PWL, and pay factor (PF) for the above authentic test results (combined) are as follows:

- Sample Mean = 11.253 in
- Sample Standard Deviation = 0.388 in
- Lower Quality Index (QIL) = 0.651
- PWL = 73.75%
- Authentic PF = 91.88%

The above pay factor (91.88 percent) can be increased by manipulating the test results to either increase the mean value or reduce the standard deviation. The data was manipulated using these approaches (using the process discussed earlier in Case 1). The results are shown in Tables 3.7 and 3.8

The QC data for PCCP thickness can be manipulated by decreasing the sample standard deviation to 0.177 in without being detected by the F-test (see Table 3.7). When the sample standard deviation was decreased to 0.157 in, however, the F-test detected this manipulation. Similarly, the data can be manipulated by increasing the sample mean up to 11.80 in without being detected by the t-test (see Table 3.8). When the mean was increased to 11.84 in, however, the t-test detected this manipulation.

Table 3.7: Original and Manipulated PCCP Slab Thickness Data (by Decreasing SD)

Test No.	Original Test Results	Manipulated Data				
	SD=0.405 in	Case 1 SD= 0.352 in	Case 2 SD =0.239 in	Case 3 SD= 0.206 in	Case 4 SD = 0.177 in	Case 5 SD =0.157 in
1	11.27	10.64	11.33	11.07	11.22	11.18
2	10.68	11.49	11.41	10.98	11.15	11.32
3	11.49	11.81	11.20	11.41	11.35	11.39
4	10.81	11.22	11.60	11.28	11.30	11.02
5	11.07	11.46	11.43	11.17	11.37	11.20
6	11.18	11.13	11.14	11.33	10.97	11.29
7	11.91	10.99	11.00	11.19	11.52	11.52
8	11.58	11.31	10.87	11.64	11.09	11.13
Sample Mean	11.250	11.259	11.246	11.258	11.245	11.257
Sample Std. Dev	0.405	0.352	0.239	0.206	0.177	0.157
p-value for F-test	0.868	0.655	0.214	0.119	0.060	0.034
Reject H_0 for F-test?	No	No	No	No	No	Yes
p-value for t-test	0.972	1.000	0.947	0.996	0.937	0.992
Reject H_0 for t-test?	No	No	No	No	No	No
Quality Index	0.651	0.732	0.871	0.954	0.971	NA
PWL	73.75	76.3	80.5	82.75	83.25	NA
Pay Factor	91.875	93.15	95.25	96.375	96.625	NA

SD: Standard Deviation, NA: Not Applicable (difference detected)

Table 3.8: Original and Manipulated PCCP Slab Thickness Data (by Increasing the Mean Value)

Test No.	Original Test Results	Manipulated Data				
	Mean=11.262 in	Case 1 Mean=11.336 in	Case 2 Mean=11.560 in	Case 3 Mean=11.646 in	Case 4 Mean=11.795 in	Case 5 Mean=11.836 in
1	11.27	11.44	11.66	11.72	11.88	11.89
2	10.68	11.33	11.59	11.56	11.76	11.78
3	11.49	11.59	11.42	11.81	12.01	12.00
4	10.81	11.87	12.19	12.21	12.43	12.52
5	11.07	11.13	11.71	11.48	11.64	11.57
6	11.18	11.69	11.74	12.07	12.11	12.19
7	11.91	11.07	11.20	11.29	11.48	11.43
8	11.58	10.57	10.97	11.03	11.05	11.31
Sample Mean	11.250	11.336	11.560	11.646	11.795	11.836
Sample Std. Dev	0.405	0.412	0.370	0.391	0.420	0.403
p-value for F-test	0.868	0.893	0.732	0.816	0.925	0.862
Reject H_0 for F-test?	No	No	No	No	No	No
p-value for t-test	0.972	0.765	0.228	0.142	0.062	0.042
Reject H_0 for t-test?	No	No	No	No	No	Yes
Quality Index	0.651	0.787	1.165	1.219	1.290	NA
PWL	73.75	78.3	88	89.25	90.5	NA
Pay Factor	**91.875**	**94.15**	**99**	**99.625**	**100.25**	NA

**NA: Not Applicable (difference detected)*

3.2. Probability Profile

The above examples give an idea about how the statistical tests fail to detect the difference at certain values of standard deviation and mean. The data analyzed in the example was randomly generated, so there is a chance that this might be a random event for which the statistical tests failed to detect the difference between manipulated data and the original data while it might detect the difference for some other data sets. To overcome this possibility, 1000 data sets for each case were generated and the probability of the statistical test detecting the difference at certain value of standard deviation or mean for different sample size ratio was determined. This probability was computed as follows:

$$\text{Probability of detecting difference} = \frac{\text{No.of difference detected}}{Total\ No.of\ Simulations} \qquad \text{... Eq. 3.1}$$

For example, 1000 different data sets at every manipulated standard deviation value were generated for the case discussed earlier in section 3.1.1 to find out the probability of detecting the difference at every value of standard deviation. As discussed in Section 3.1.1, the AQC is asphalt content, the agency to contractor sample size ratio is 1:10, and the original standard deviation is 0.22%. As shown in Table 3.9, the probability of detecting a difference between the agency's data and the contractor's manipulated data remains relatively low even if the contractor's sample standard deviation was reduced from 0.22% to 0.10%.

Table 3.9: Probability of Detecting Difference in Standard Deviation.

Manipulated Standard Deviation	Ratio of Manipulated SD to Original SD	Probability of Detecting Difference
0.20	0.91	0
0.18	0.81	0
0.15	0.68	0
0.12	0.55	0.073
0.10	0.45	0.368

Similar analysis was carried out for all other sample size ratios (i.e. keeping the agency's sample size constant and varying the contractor's sample size, n_c). Figure 3.2 to Figure 3.5 show the graphical representation of the results obtained for various sample size ratios. It was observed that the probability of detecting the difference increases with

- Increase in contractor's sample size.
- Increase in difference between original values and the manipulated sample values.

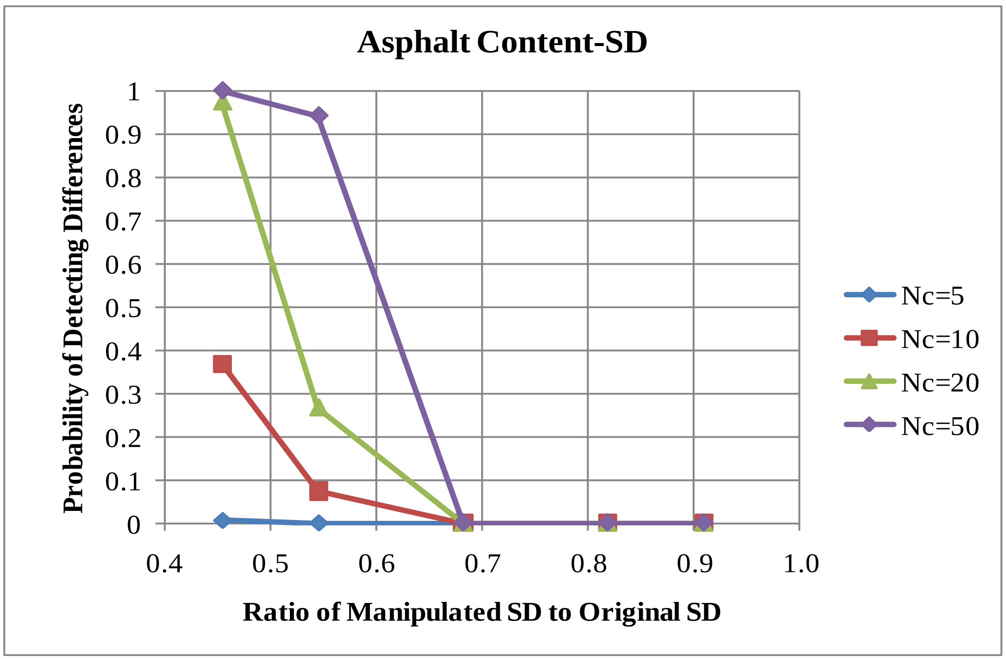

Figure 3.2: Probability of Detecting Reduced Standard Deviation of Asphalt Content

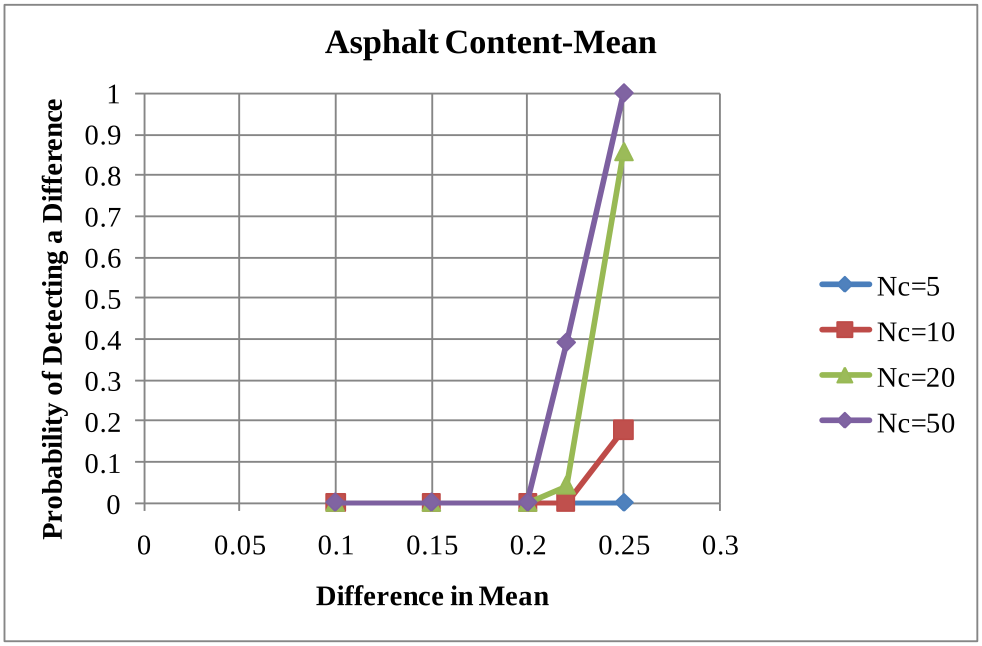

Figure 3.3: Probability of Detecting Increase in Mean of Asphalt Content

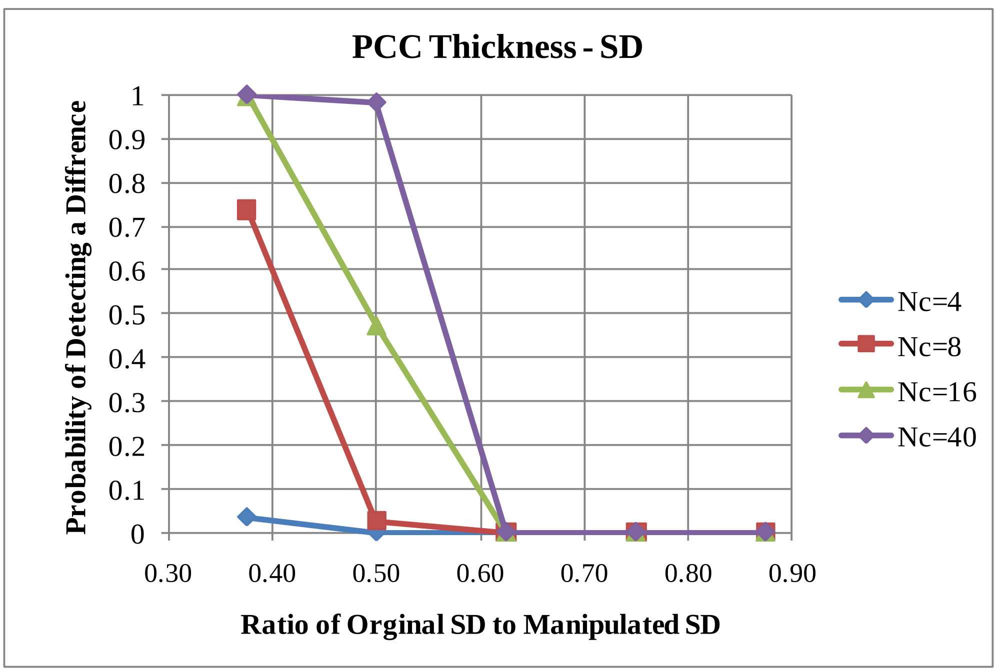

Figure 3.4: Probability of Detecting Reduced Standard Deviation of PCC Thickness

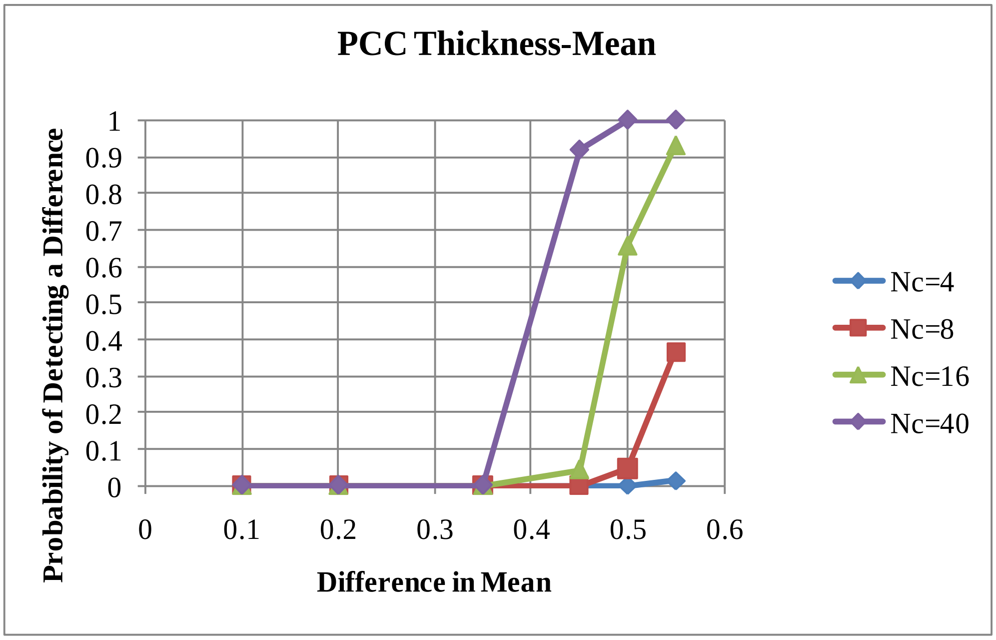

Figure 3.5: Probability of Detecting Increase in Mean of PCC Thickness

3.3. Effect of Contractor's Sample Size

The effect of contractor's sample size on the ability of the F-test and the t-test to detect manipulated sample data was analyzed. It was noted that as the contractor's sample size increases, the ability of manipulating sample data without being detected decreases. Tables 3.10 to 3.13 show the results of this analysis. For example, a contractor taking 50 samples of asphalt content would be able to reduce the standard deviation from 0.22% to 0.135% without being detected by the F-test; whereas a contractor taking 10 samples would be able to reduce the sample's standard deviation from 0.22% to 0.103% without being detected by the F-test. Another example is a contractor taking 50 samples of PCCP slab thickness would be able to increase the mean from 11.25in to 11.652in, without being detected by the t-test; whereas a contractor taking 10 samples would be able to increase the mean from 11.25 in to 11.795 in, without being detected by the t-test.

Even though the statistical tests detect the manipulation in the data (as explained above), the magnitude of increase in pay factor from the original value increases with increase in contractor's sample size. This trend was observed for both asphalt content and PCC thickness. Also, the magnitude of increase in pay factor for a given sample size is greater if manipulation is carried out by adjusting the mean.

Table 3.10: Maximum Undetected Decrease in Standard Deviation in Asphalt Content

Sample Size		Contractor's Original Sample		Contractor's Manipulated Sample		Pay factor		Increase in Pay Factor
Agency	Contractor	Mean	Std. Dev.	Mean	Std. Dev.	Original	Manipulated	
5	5	4.910	0.212	4.905	0.076	88	90.165	2.165
5	10	4.902	0.213	4.903	0.105	87.65	90.5	2.85
5	20	4.907	0.220	4.901	0.119	87.65	90.835	3.185
5	50	4.901	0.217	4.902	0.135	87.55	90.95	3.4

Table 3.11: Maximum Undetected Increase in Mean in Asphalt Content

Sample Size		Contractor's Original Sample		Contractor's Manipulated Sample		Pay factor		Increase in Pay Factor
Agency	Contractor	Mean	Std. Dev.	Mean	Std. Dev.	Original	Manipulated	
5	5	4.904	0.226	5.210	0.226	88	95.215	7.215
5	10	4.902	0.213	5.149	0.228	87.65	96.55	8.9
5	20	4.904	0.217	5.120	0.218	87.65	98.4	10.75
5	50	4.903	0.227	5.117	0.228	87.5	99.025	11.525

Table 3.12: Maximum Undetected Decrease in Standard Deviation in PCC Thickness

Sample Size		Contractor's Original Sample		Contractor's Manipulated Sample		Pay factor		Increase in Pay Factor
Agency	Contractor	Mean	Std. Dev.	Mean	Std. Dev.	Original	Manipulated	
4	4	11.25	0.400	11.250	0.102	92.5	96.5	4
4	8	11.25	0.405	11.245	0.177	91.875	96.625	4.75
4	16	11.25	0.423	11.252	0.209	91.65	97.25	5.6
4	40	11.25	0.406	11.250	0.232	91.835	97.3	5.465

Table 3.13: Maximum Undetected Increase in Mean in PCC Thickness

Sample Size		Contractor's Original Sample		Contractor's Manipulated Sample		Pay factor		Increase in Pay Factor
Agency	Contractor	Mean	Std. Dev.	Mean	Std. Dev.	Original	Manipulated	
4	4	11.25	0.399	11.907	0.399	92.375	98.75	6.375
4	8	11.25	0.405	11.795	0.419	91.875	100.25	8.375
4	16	11.25	0.413	11.712	0.406	91.65	101.25	9.6
4	40	11.24	0.409	11.652	0.414	91.165	101.45	10.285

3.4. Conceptual Pitfalls: Intermingling Process Control and Product Acceptance

One of the pillars of modern quality control theory (as illustrated in Deming's 14 tenets) is the focus on defect prevention through process control (not defect detection and containment through mass inspection) *(12)*. This requires the contractor to focus on "process control" tests (not "product acceptance" tests). Thus the use of contractor test results for acceptance purposes contradicts the principles of quality control theory.

According to quality control theory, the purpose of quality control tests is to identify quality problems during materials production and construction so that adjustments can be made to maintain desirable quality level; while the purpose of acceptance tests is to estimate the quality of the delivered product so that acceptance and pay adjustment decisions can be made accordingly. This approach to quality control and product acceptance is depicted in the models shown in Figures 3.6 and 3.7. In these models, the contractor should focus on “process control” to identify and ultimately remove the underlying causes of the problem (i.e., prevention rather than identification and containment of defective material) *(13)*. Thus, process control data collection (including testing) should occur as early as possible in the process. Acceptance testing, on the other hand, should occur as late as possible in the process (to be as representative as possible of the final in-service product).

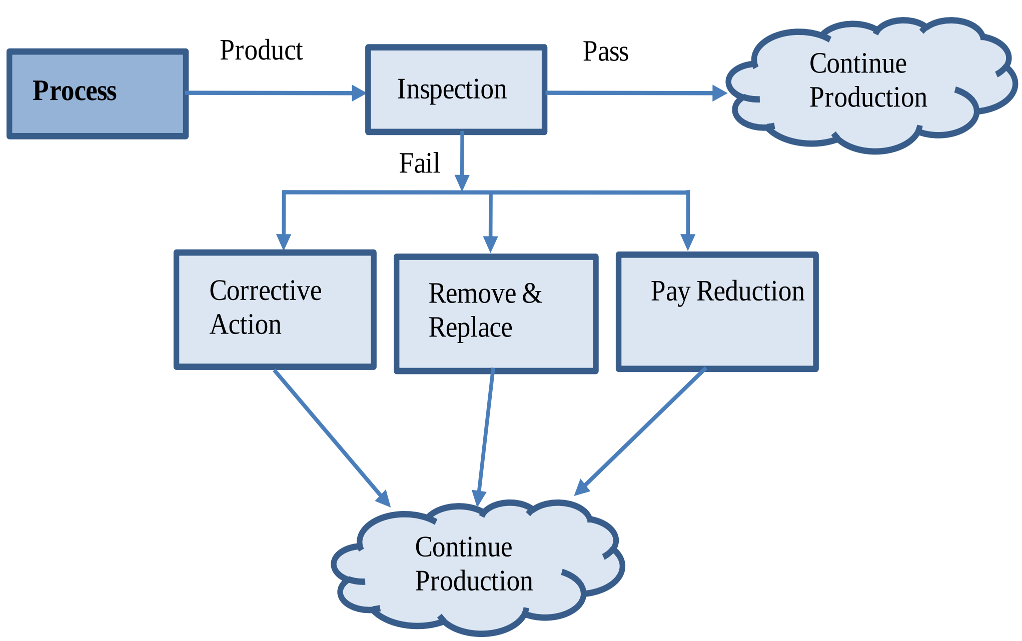

Figure 3.6: Product-Focused Model for Construction and Materials Acceptance

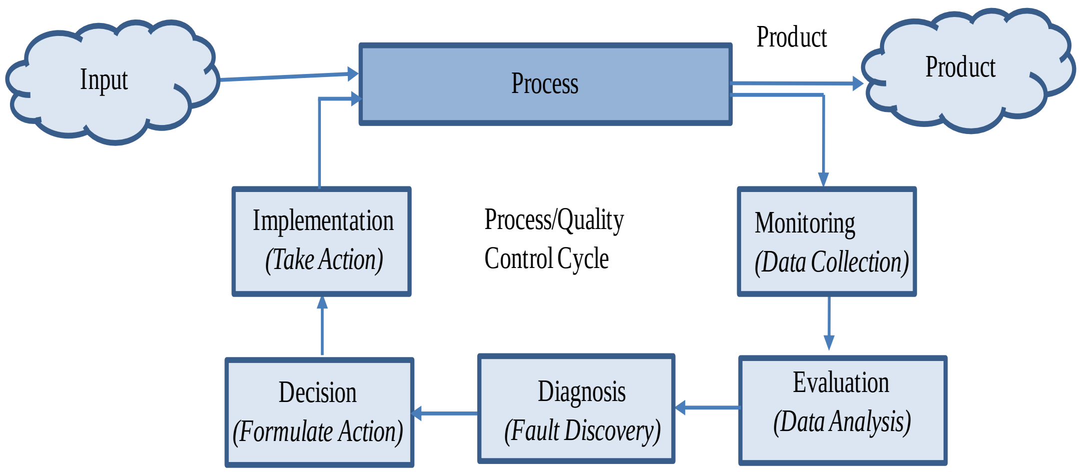

Figure 3.7: Process-Focused Model for Construction and Materials Quality Control

(Figures 3.6 and 3.7 are adopted from *(13)*)

The highway construction and material quality assurance literature recognizes this approach to quality by identifying product acceptance, quality control, and independent assurance as three separate functions of quality assurance [see *(6),(9) for definition*]. Additionally, the use contractor's quality control data for acceptance decisions encourages mere conformance to specification limits, and thus provides less emphasis on uniformity in the production process. This is illustrated in Figures 3.8 and 3.9. Figure 3.8 shows an acceptance-oriented process that lacks uniformity. Figure 3.9 shows a quality- and uniformity-oriented process where consistent results are obtained after an error was identified and corrected (i.e., the process was brought to control).

Finally, it should be noted that quality control theory does not preclude the contractor from using acceptance test results (performed by the agency) to help in the process control for subsequent lots.

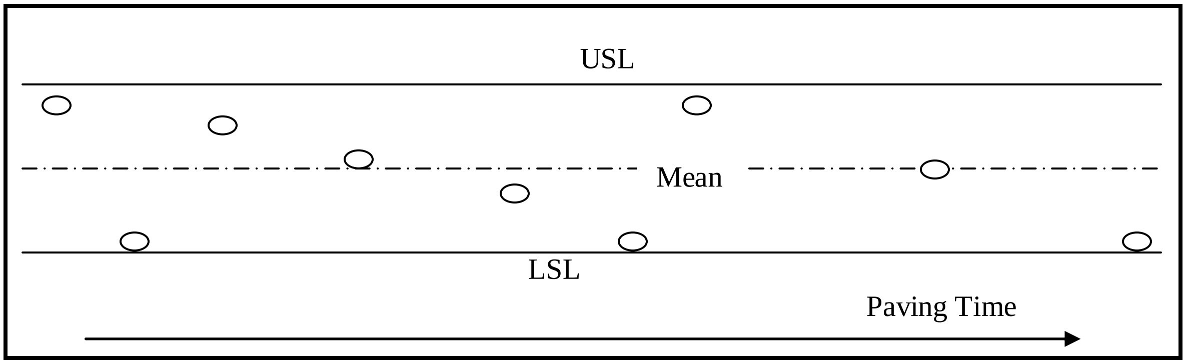

Figure 3.8: Expected Outcome of Product Acceptance

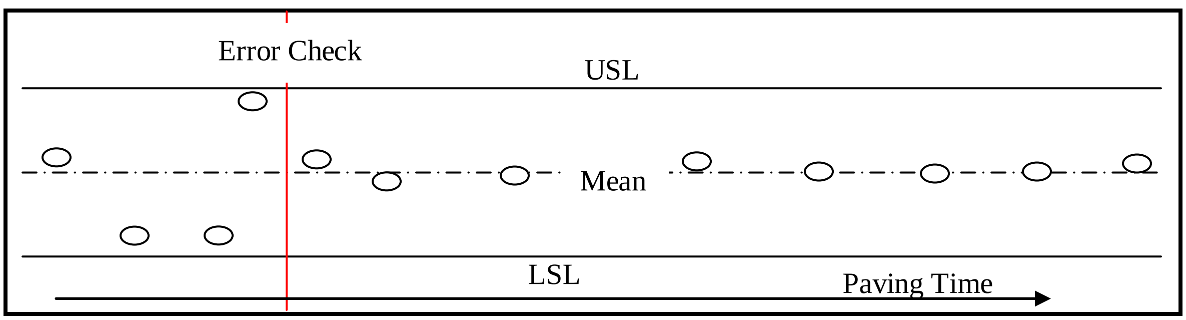

Figure 3.9: Expected Outcome of Process Control

4. ALTERNATIVES AND IMPROVEMENTS TO CONTRACTOR ACCEPTANCE TESTING

This chapter discusses the results of a workshop that was held in 2009 at the FHWA to identify and evaluate potential alternatives and improvements to the use of contractor test results for acceptance purposes.

4.1. Workshop Overview

The workshop was held at the Turner-Fairbank Highway Research Center (TFHRC) in McLean, VA, on February 2, 2009. Attendances included 10 technical working panel (TWP) members from state DOTs, paving industry, consultants, and academia, and two non-members from the FHWA. This workshop was regarded as a "brainstorming" session, in which the participants discussed, evaluated advantages and disadvantages of, and subsequently ranked different alternatives and improvements to the use of contractor test results for acceptance purposes.

An initial set of 12 potential alternatives and improvements were proposed to the TWP members. Discussions and comments were made on these alternatives and improvements. The results of these discussions are introduced in the following sections.

4.2. Alternatives and Improvements to Contractor Acceptance Testing

A set of alternatives and improvements to the use of contractor's test results for acceptance purposes was developed based on a review of the literature. These alternatives and improvements were grouped into four categories as shown in Table 4.1.

Table 4.1: Initial Set of Alternatives and Improvements to Contractor Acceptance Testing

Category	Alternative/Improvement
- Alternatives aimed at reducing amount/frequency of agency testing	− Start project with normal testing frequency and, then reduce frequency (i.e., increase lot size or reduce sample size) once there is evidence that the contractor's process is under control. − Reduce testing of each AQC and randomize the AQCs to be tested at any one location. − Reduce sample size to 3 per lot. − Reduce or eliminate the averaging of multiple (i.e., replicate) samples.
- Alternatives aimed at delegating acceptance test	− Use third-party testing for acceptance purposes (e.g., commercial lab representing the agency). − Use of automated equipment and plant records.
- Alternatives that use contractor qualifications	− Test contractors with "A" ratings at lower frequency than contractors with "C" ratings in conjunction with o Stronger independent assurance program to prevent abuse; o Post construction evaluations of contractors. − Require certain certification and/or training of the contractor's technicians.
- Potential improvements to contractor acceptance testing	− Eliminate or reduce bonuses to decrease the potential for fraud. − Use larger lots to compare contractor vs. agency test results (F- and t- tests would have larger n and be more discerning). − Use contractor's QC data in acceptance decisions. − Combine contractor and agency test results.

The TWP members discussed, and then evaluated and ranked these alternatives. Additional potential alternatives were identified during the discussion. The following subsections summarize the TWP discussions of these alternatives and improvements.

Alternative 1.1- Start project with normal testing frequency and then reduce the frequency (i.e., increase lot size or reduce sample size) once there is evidence that the contractor's process is under control.

It should be noted that if quality of production shows signs of degradation, the agency needs to revert back to high frequency tests. This approach has been used by Florida DOT (FDOT). Indiana DOT (INDOT) is considering using this technique (called "risk-based" inspection). Positive comments included that this alternative might reduce cost of testing to the agency and that it can weed out the quality-oriented contractors from poor-quality contractors (i.e., those who do not place as much importance on quality). However, some contractors held a negative opinion on this alternative as they thought it increases project uncertainty and thus may result in higher bids. Some members of the TWP suggested that this alternative may be difficult to administer. A formalized version (Skip-lot Sampling) of this alternative is discussed later in Section 5 of this thesis.

Alternative 1.2 - Reduce testing of each AQC and randomize the AQCs to be tested at any one location

No positive comments were rendered on this alternative. This alternative was commonly thought to be difficult to administer. Developing and implementing a statistically sound acceptance plan with varying sample size (*n*) and multiple randomized AQCs is a complex task for most DOTs.

Alternative 1.3 - Reduce sample size to 3 per lot

Economic analysis of sample size *(14)* shows that a sample size of 3 is most economic to the agency, for most practical cases. However, no general agreement among the TWP members was found on this alternative. TWP members indicated that this alternative could be resisted by both good contractors and poor contractors (as good contractors want their quality to be accurately estimated while poor contractors want the DOT test results to help them with process control). However, reduced sample size, can potentially be effective if linked to project criticality (e.g., as measured by traffic level or highway classification), so that the sample size on non-critical projects (e.g., non-Interstate Highways or low traffic roads) can be reduced.

Alternative 1.4 - Reduce or eliminate the averaging of multiple (i.e., replicate) samples

It was suggested that, from a practical viewpoint, replicates are needed to account for outliers in test results.

Alternative 2.1 - Use third-party testing for acceptance purposes (e.g., commercial lab representing the agency).

Virginia DOT uses this method. TWP members suggested that this method may increase the cost of sampling and testing for the DOT. And, this alternative may not be effective in fighting the potential for data manipulation.

Alternative 2.2 - Use of automated equipment and plant records to replace/decrease testing of asphalt content, gradation, air content, strength, etc.

Some material production plants have already gone through vigorous quality programs. However, several potential disadvantages were noted, such as plant records may not reflect field (as-built) quality; it may lead to less QC testing; equipment needs regular calibration, and equipment records are normally limited. Additionally, workmanship-related deficiencies might be difficult to detect with automated equipment.

Alternative 3.1 - Test contractors with "A" rating at lower frequency than contractors with "C" rating, in conjunction with a) Stronger independent assurance program to prevent abuse, and b) Post construction evaluations of contractors.

FDOT has established a contractor grading system that defines what projects a contractor can bid on. This alternative was believed to be able to encourage poor contractors to step up. Amount of testing could be reduced since "A" rated contractors could be tested less or not at all. A flat fee for acceptance testing can be assessed. This fee (or a portion of it) can be passed on to the well-rated contractor as an incentive, if the state is not required to perform as much testing. Negative opinions on this alternative included that ratings may vary from state to state and it is hard for a contractor to bid on projects in a state where they have not set any history yet to get good ratings. The cost to administer this alternative may be very high unless the state already has some prequalification program in place.

Alternative 3.2 - Require certain certification and/or training of the contractor's technicians

FDOT and INDOT are using this approach on their projects. No more comments were made on this alternative.

Improvement 4.1 - Eliminate or reduce bonuses to decrease the potential for fraud

There was no positive support for this alternative. TWP members suggested that if bonuses are reduced or eliminated, pay reductions should also be reduced or eliminated.. Also, if bonuses are reduced, contractors would have less incentive to achieve higher quality because the cost to get the bonus may outweigh the actual bonus. If bonuses are eliminated and disincentives remained, in the long run, the contractor would not achieve an expected pay of 100%.

Improvement 4.2 - Separate the contractor's testing staff from the contractor's project management staff

This approach requires contractor's testing staff to report to a separate unit within the contractor's organization. This can potentially relieve the contractor's testing staff from possible pressures from the project managers to produce favorable test results. Thus, it can potentially help fight fraud.

Improvement 4.3 - Use larger lots to compare contractor test results to agency test results; F- and t- tests would have larger n and be more discerning

It was pointed out that a hot-mix asphalt (HMA) project must be at least 10,000 tons to generate sufficient sample units for reliably verifying the contractor test results using F- and t- tests. This argument supports larger lots as an improvement to the practice of using contractor acceptance testing. TWP members noted that with larger lots, a) the normality of data obtained from larger lot sizes should be statistically checked because the F and t-tests assume that the data come from a normal distribution and b) DOT should consider linking increased lot size (and thus reduced testing frequency) to project criticality (e.g., as measured by traffic level or highway classification), so that larger lots are used on non-critical projects (e.g., non-Interstate Highways or low-traffic roads).

Additional alternatives and improvements

The TWP members identified the following additional alternatives and improvements:

- Use warranties
- Slow the project down to give agency more time to run tests
- Require certain certification and/or training of the contractor's technicians who perform acceptance testing
- Develop guidelines for applying F and t tests for contractor acceptance testing
- Make no changes to current practices.

4.3. Evaluation Method of the Eighteen Potential Alternatives/Improvements

Subsequent to the workshop, five members of the TWP (from both the industry and government agencies) evaluated the above alternatives and improvements on the basis of three main criteria, which are shown in Table 4.2.

Table 4.2: Evaluation Criteria of Identified Alternatives

No.	Criteria	Description
1	Potential for Reducing Agency's Workload	How much of the current workload can be reduced by adopting a certain alternative
2	Potential for Increasing Agency's Risk of Accepting Poor Quality Products	What is the probability that if a certain alternative is adopted, it would make the agency more vulnerable to fraud or low quality material
3	Ease of Implementation	How easy it would be for the agency to implement the alternative in the field considering organizational, economical, and political realities of highway construction projects

Each criterion in Table 4.2 had three descriptive rating levels: Low, Medium, and High. The evaluators were asked to use these levels to rate each alternative/improvement according to each criterion in Table 4.2. These rating levels were then converted to a numerical scale to facilitate the ranking of all identified alternatives/improvements. For Criteria # 1 and 3 (where High is desirable), a score of 3 was assigned to the High rating, 2 assigned to the Medium rating, and 1 assigned to the Low rating. For Criterion # 2 (where High is undesirable), the numerical scoring was done in the reverse way: 3 assigned to the Low rating, 2 assigned to the Medium rating, and 1 assigned to the High rating. It should be noted that some evaluators chose the mid (or combined) ratings of Low-Medium and Medium-High. In these cases, for Criteria # 1 and 3, a score of 1.5 was assigned to the Low-Medium rating; and a score of 2.5 was assigned to the Medium-High rating. These scores were reversed for Criterion # 2. The responses given by the panel members are provided in Appendix B.

For each alternative, an average score for each criterion was computed by dividing the sum of all the points (from the five respondents) by five. The three criteria were regarded as equally important. Thus the overall average score for each alternative was determined by dividing the sum of the scores for all the three criteria by three.

An additional question (i.e., whether an alternative deserves further investigation) was also asked in the evaluation form. It was a multiple-choice problem with the options of “Yes”, “No”, and “Maybe.” To score the alternatives/improvements based on this additional question, a score of 1 was given to a “No” answer, 2 was given to a “Maybe” answer, and 3 was given to a “Yes” answer. The average score for each alternative/improvement was computed. Finally, the alternatives/improvements were ranked to determine their worthiness of further investigation based the average score.

4.4. Results of the Evaluation

Based on the scoring method discussed in the previous section, the studied improvements/alternatives were ranked according to:

- Overall average score (considering all three evaluation criteria)
- Average score for Criterion # 1 (Potential for Reducing Agency's Workload)
- Average score for Criterion # 2 (Potential for Increasing Agency's Risk of Accepting Poor Quality Products)
- Average score for Criterion # 3 (Ease of Implementation)
- Average score for Worthiness of Further Investigation

The top five alternatives/improvements according to the above rakings are presented in Tables 4.3 through 4.7. The scores for all alternatives/improvements are presented in Appendix B.

Table 4.3 shows the top five alternatives based on overall rating of the three evaluation criteria. The "*Use warranties"* alternative ranked the first with an average score of 2.7. Followed are the options of third-party testing, larger lot sizes, making no changes to current practices, and automated equipment and plant records.

Table 4.3: Top 5 Alternatives Based on Overall Average Score

Ranking	Alternatives	Average Score (out of 3.0)
1	Use warranties.	2.7
2	Use third–party testing for acceptance (e.g. by commercial lab representing the agency).	2.57
3	Use larger lot sizes.	2.43
4	Make no changes to current practices.	2.33
5	Use automated equipment and plant records to replace/decrease testing of asphalt content, gradation, air content, strength, etc.	2.28

Table 4.4 presents the top five alternatives considering the first criterion only (Potential for Reducing Agency's Workload). The alternative "*Use third-party testing for acceptance"* was ranked as the first choice for reducing agency's workload. Four options had equal average score and thus were tied in the fifth position.

Table 4.4: Top 5 Alternatives Based on Criterion #1 (Potential for Reducing Agency's Workload)

Ranking	Alternatives	Average Score (out of 3.0)
1	Use third–party testing for acceptance (e.g. by commercial lab representing the agency).	2.7
2	Use warranties.	2.6
3	Use larger lot sizes.	2.5
4	Test contractors with "A" ratings at a lower frequency than contractors with "C" ratings. Contractor ratings are for quality management purposes only, with no effect on bidding.	2.2
5a	Use automated equipment and plant records to replace/decrease testing of asphalt content, gradation, air content, strength, etc.	2.0
5b	Reduce sample size to 3 per lot.	2.0
5c	Randomize the AQCs to be tested at any one location (i.e., do not test all AQCs at all locations).	2.0
5d	Combine contractor and agency test results.	2.0

Table 4.5 shows the results for the second criteria (Potential for Increasing Agency's Risk of Accepting Poor Quality Products). Four alternatives/improvements were ranked in the first place, with an equal score of 3.0 (out of 3.0).

Table 4.5: Top 5 Alternatives Based on Criterion #2 (Potential for Increasing Agency's Risk of Accepting Poor Quality Products)

Ranking	Alternatives	Average Score (out of 3.0)
1	Require certain certification and/or training of the contractor's technicians who perform acceptance testing.	3.0
2	Use larger lots to compare contractor vs. agency test results; F and t tests would have larger n and thus be more discerning (conditioned on normality of data).	3.0
3	Require contractor's testing staff to report to a separate unit within the contractor's organization (i.e., require a separation between the contractor's quality management team and project management team).	3.0
4	Make no changes to current CAT practices.	3.0
5	Slow the project down to give agency more time to run tests.	2.7

Table 4.6 shows the top alternatives considering the third criterion only (Ease of Implementation). Use larger lot sizes, Make no changes to current practices, and Reduce sample size to 3 per lot were tied in the first place, with an equal score of 3.0

Table 4.6: Top 5 Alternatives Based on Criterion #3 (Ease of Implementation)

Ranking	Alternatives	Average Score (out of 3.0)
1a	Use larger lot sizes.	3.0
1b	Make no changes to current CAT practices.	3.0
1c	Reduce sample size to 3 per lot.	3.0
2a	Use warranties.	2.8
2b	Use automated equipment and plant records to replace/decrease testing of asphalt content, gradation, air content, strength, etc.	2.8
2c	Use third–party testing for acceptance (e.g. by commercial lab representing the agency).	2.8

Table 4.7 shows the top five alternatives worthy of further investigation. The alternatives of using warranties, certification of contractor's technicians, larger lot sizes, separation of contractor's testing staff from project management, and automated equipments were considered deserving further investigation than the other alternatives.

Table 4.7: Top 5 Alternatives Based on the Worthiness of Further Study

Ranking	Alternatives	Average Score (out of 3.0)
1	Use warranties.	3.0
2	Require certain certification and/or training of the contractor's technicians who perform acceptance testing.	3.0
3	Use larger lots to compare contractor vs. agency test results; F and t tests would have larger n and thus be more discerning (conditioned on normality of data).	3.0
4	Require contractor's testing staff to report to a separate unit within the contractor's organization (i.e., require a separation between the contractor's quality management team and project management team).	3.0
5	Use automated equipment and plant records to replace/decrease testing of asphalt content, gradation, air content, strength, etc.	2.8

5. SKIP-LOT SAMPLING PLANS

The concepts and procedures of Skip-Lot Sampling Plans (SkSPs) as a method for reduced sampling and testing workload are introduced in this section. Skip-lot sampling is studied here as a formal acceptance method for implementing alternatives 1.1 (reduced sampling frequency) and 4.3 (larger lot size), discussed in Section 4 of this report. The application of SkSP to highway construction and materials quality assurance is illustrated through an example problem. SkSP was identified as a potential alternative to contractor's acceptance testing subsequent to the TWP workshop; and thus was not evaluated by the TWP members.

5.1. Rationale and Background of Skip-lot Sampling Plan

Current acceptance sampling plans for highway construction and materials require sampling and testing of every individual lot (i.e., 100 percent of the lots are inspected). This is appropriate if the contractor is erratic. But, if the contractor is fairly steady, should, or can, the agency (i.e., the buyer) take that into consideration, and by doing so reduce the sampling and testing workload. This is the rationale for Skip Sampling, which was introduced by Harold F. Dodge at the Bell Telephone Laboratories in the 1950s *(15)*. Dodge introduced skip-lot sampling as a means for reducing acceptance testing by taking past quality into consideration. This technique can potentially be used for reducing sampling and testing workload required by highway acceptance plans.

SkSP went through several improvements since it was originally introduced in the 1950. The operating characteristics of Dodge's initial skip-lot sampling plan (commonly referred to as SkSP-1) were not addressed explicitly *(16)*. This limitation was later addressed by Dodge and Perry [see *(17), (18)*] and a new version of skip-lot sampling plan was developed and labeled as SkSP-2. Subsequent improvements to skip-lot sampling were made through the efforts of Parker and Kessler *(19)*. The methods of skip-lot sampling plan were eventually standardized in 1987 as *Skip-Lot Sampling Standard*, ANSI/ASQC Standard S1-1987. Currently, SkSP is used in many industries such as semiconductor manufacturing *(20)*.

SkSP is generally applicable to bulk materials or products produced or furnished in successive batches or lots. The basic conditions for applying skip-lot sampling are *(15)*:

- The product is comprised of a series of successive lots of material that come from the same source and are of essentially the same quality.
- The specification requirements are expressed as upper and/or lower limits.
- For any given AQC, the normal acceptance procedure for each lot is to obtain a suitable sample of the material and subject it to a particular test. The lot is considered *conforming* if the test results are within the specification limits, and *nonconforming* if the test results are outside specification limits.

If the acceptance decision is made based on multiple AQCs, it is not required to apply skip sampling simultaneously to all of the AQCs. Instead, it can be applied to one or more, as long as the above assumptions hold. Generally, skip sampling should be applied to those AQCs that involve the most time and labor consuming sampling and testing. If the plan is applied to multiple AQCs at the same time, it would be preferable to avoid omitting all qualified tests on some lots and performing all such tests on others. Judgment should be used in spreading the testing schedule *(15)*.

Finally, to prevent possible misuse of the plan, Dodge recommended that skipped lots be selected in a random manner. For example, if the plan calls for skipping 50 percent of the lots, a lot can be selected for testing (or skipping) by tossing a coin.

5.2. Skip-lot Sampling Plan-1 (SkSP-1)

Dodge *(15)* initially presented the skip-lot sampling plan (designated as SkSP-1) as an extension of the continuous sampling plan (CSP-1), which was designed for individual units of production. However, SkSP-1 considers a series of lots, not a series of product units.

SkSP-1 is defined by two parameters: number of successive confirming lots required to qualify for skip-lot inspection (called clearing interval, i) and the fraction of lots inspected during skip-lot sampling (called fraction, f). The process of SkSP-1 consists of the following steps *(15)*:

- Step 1: At the outset, test every lot consecutively and continue such testing until i lots in succession are found to be conforming.
- Step 2: When i lots in succession are found to be conforming, discontinue testing every lot, and instead, test only fraction f of the lots.
- Step 3: If a tested lot is found to be nonconforming:

- Either (a) require a corrective action, or (b) remove and replace the nonconforming lot by a conforming lot, and
- Revert immediately to testing every consecutive lot until again *i* lots in succession are found conforming (i.e., revert to Step 1).

Dodge *(14)* has shown that the average outgoing quality (P_A) can be computed as a function of *i*, *f*, and product's percent defective as follows:

$$P_A = p[1 - \frac{f}{f + (1-f) \times (1-p)^i}] \quad \text{…Eq. 5.1}$$

where,

p = product's percent defective

i = clearing interval (i.e., number of consecutive confirming lots required to qualify for skip-lot sampling), a positive integer.

f = fraction of lots tested during kip-lot sampling, *f* ($0 < f < 1$).

The value of *p* (in Equation # 5.1) for which the maximum value of P_A occurs is referred to as the average outgoing quality limit (AOQL) and is used to express the degree of protection a SkSP-1 can offer. For example, an AOQL value of 2 percent indicates that an average of not more than 2 percent of accepted lots will be nonconforming for the AQC under consideration. Figure 5.1 can be used to determine AOQL as a function of *i* and *f*. For example, a SkSP-1 plan with *i=14* and *f=0.5*, results in an AOQL of 2 percent. AOQL is similar in purpose to AQL in conventional sampling plans.

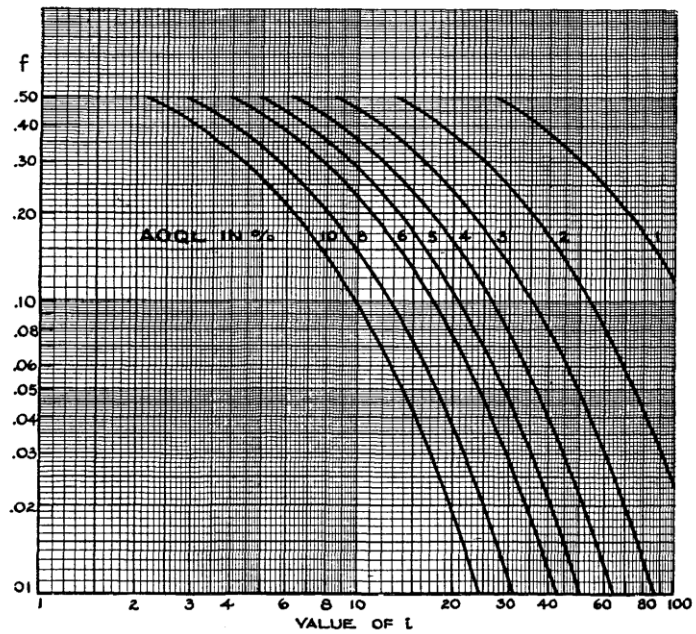

Figure 5.1: Curves for Determining Values of AOQL for Given *f* and *i, and* vice versa *(15).*

5.3. Skip-lot Sampling Plan-2 (SkSP-2)

Dodge and Perry [*(17), (18)*] extended SkSP-1 to a system of sampling by incorporating a "reference sample plan" for accepting or rejecting each lot. While, SkSP-1 did not preclude the use of a lot-by-lot acceptance sampling plan for assessing the conformance of each tested lot, the operating characteristics for such combination were not explicitly addressed *(16)*. Perry *(18)* proposed the next logical step in SkSP-2; where each lot to be inspected is sampled according to some attribute (with possible extension to variable) lot-inspection plan *(16)*. This lot-by-lot acceptance sampling plan is called "reference sample plan." Thus, a skip-lot plan of type SkSP-2 can be described as one that uses a "reference sampling plan" for lot-by-lot acceptance together with the SkSP-1 process. Similar to SkSP-1, a SkSP-2 plan is defined by *f* (fraction of lots tested during skip-lot sampling) and *i* [clearing interval (i.e., number of consecutive confirming lots required to qualify for skip-lot sampling]; where *i* is a positive integer and *f* ($0 < f < 1$).

For highway projects, the skip-lot plan SkSP-2 can be graphically depicted as shown in Figure 5.2. In this sketch, "A_t" is accepted lot under the reference plan; "R" is rejected lot under the reference plan; "A_s" is accepted lot due to skipping (i.e., lot is accepted without testing); "*U*" is the expected number of lots during "normal inspection" (also known as "qualification inspection"); and "*V*" is the expected number of lots during "skipping inspection," until reverting back to testing every consecutive lot. During qualification inspection, every lot is sampled and tested using the reference plan. During skipping inspection, lots are skipped and only a fraction *f* of the total lots is selected for sampling and testing.

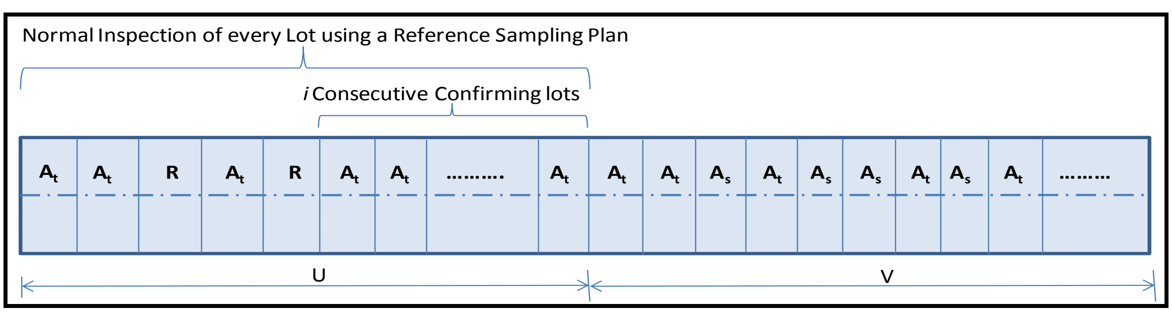

Figure 5.2: A Sketch of a SkSP-2 Plan for Highway Construction and Materials Lots.

Perry *(18)* developed the concept of "operating ratio" (OR) to help select the skipping parameters for SkSP-2. According to Perry *(18)*, OR is computed as follows:

$$OR = \frac{P_{10}}{P_{95}} \quad \text{...Eq.5. 2}$$

where,

P_{10} = product's percent defective to which the work should receive a 10% probability of acceptance

P_{95} = product's percent defective to which the work should receive a 95% probability of acceptance

In conventional acceptance plans for highway construction and materials, P_{10} and P_{95} can be viewed as the equivalents of rejectable quality limit (RQL) and acceptable quality limit (AQL), respectively. OR reflects the ability of the acceptance plan to discriminate between good and bad quality. Dodge and Perry *(16)* developed tables that can be used to select adequate combinations of *f* and *i* values for any given OR and attribute reference sampling plan (as expressed in the acceptance number, *c*). These tables are provided in Appendix C of this report.

Perry *(21)* used a power series approach and a Markov chain technique to develop operating characteristics of SkSP-2 plans. Let *P* denote the probability of accepting a lot according to the reference plan and P_a denote the corresponding probability of acceptance for the SkSP-2 plan. The operating characteristics of SkSP-2 can be computed as follows:

The average (i.e., expected) number of lots inspected (i.e. sampled) during the "qualification inspection" phase (U):

$$U = \frac{1-P^i}{P^i(1-P)} \qquad \text{... Eq. 5.3}$$

The average number of lots inspected during the "skipping inspection" phase (V):

$$V = \frac{1}{f(1-P)} \qquad \text{... Eq. 5.4}$$

The average fraction of all submitted lots that is inspected (during both "qualification inspection" and "skipping inspection" phases) (F):

$$F = \frac{f}{(1-f)P^i+f} \qquad \text{... Eq. 5.5}$$

The probability of acceptance for the SkSP-2 plan (P_a):

$$P_a = PF + (1-F) \qquad \text{... Eq. 5.6}$$

Since skipped lots have a 100 percent probability of acceptance, P_a becomes:

$$Pa = \frac{(1-f)P^i+fP}{(1-f)P^i+f} \qquad \text{... Eq. 5.7}$$

Perry *(21)* has shown that P_a is a decreasing function of f and i, but is an increasing function of P (see the figure on page 74).

The increased probability of accepting a nonconforming lot (i.e., lot that should be rejected according to the reference plan, but is accepted due to the use of skipping), is referred to as the average outgoing quality (AOQ_2) and is computed as:

$$AOQ_2 = P_a - P \qquad \text{... Eq. 5.8}$$

The average sample number (ASN) (i.e., average number of sample units inspected per lot) is computed as:

$$ASN(SkSP) = ASN(R) \times F \qquad \text{... Eq. 5.9}$$

where,

ASN(R) = average sample number of the reference sampling plan. For single sampling plans (normally used for acceptance of highway construction and materials) with a sample size of n, ASN = n, and thus:

$$ASN(SkSP) = n \times F \qquad \text{... Eq. 5.10}$$

Since F is a fraction (between 0 and 1), Equations 5.9 and 5.10 show that a skip-lot sampling plan yields a reduction in inspection of successive lots of good quality, compared to the conventional reference sampling plan. For low percent defective (i.e., high quality), a small value of f (such as 1/4 or 1/5) can be used, resulting in substantial reduction in ASN (i.e., average sample size) *(18)*. This is demonstrated through the numerical example shown in the following section of this report.

5.4. An Example Application of SkSP-2

An example problem is presented here to better understanding of the potential application of SkSP-2 to the quality assurance process of highway constructions.

Suppose that the acceptance plan for a given AQC uses percent within Limit (PWL) as the quality measure with an acceptance limit (M) of 60 percent within limit and a sample size (n) of 5. To be consistent with the literature on SkSP-2, percent defective (PD) and acceptance constant (k) are used instead of PWL and acceptance limit (M), respectively. An M of 60 PWL was converted to an equivalent acceptance constant (k) of 0.282 using statistical tables provided in the AAHTO R 9-90 *Standard Recommended Practice for Acceptance Sampling Plans for Highway Construction (21)*. The OC curve for this acceptance plan (see Figure 5.3) was constructed using statistical tables provided in the AAHTO R 9-90 *Standard Recommended Practice for Acceptance Sampling Plans for Highway Construction (22)*.

Suppose that the sate DOT typically achieves a PD of five percent defective on its projects. The following analysis shows how a SkSP-2 plan can affect the amount of required acceptance testing and the agency's buyer's risk (β).

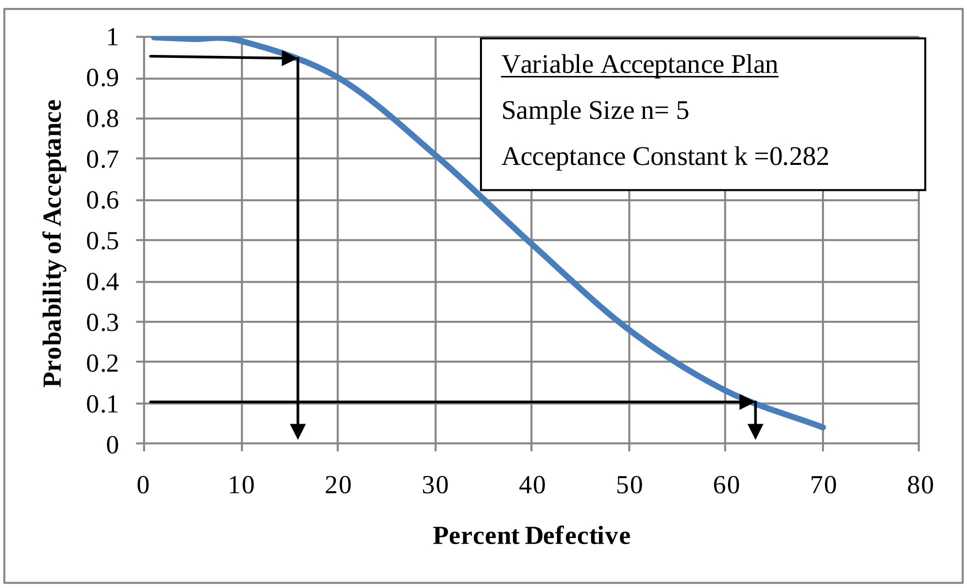

Figure 5.3: Operating Characteristics Curve for the Original Acceptance Plan

Step 1: Selection of Skipping Parameters and a Reference Sampling Plan

From the OC Curve in Figure 5.3, it can be seen that a 15 percent defective corresponds to a 95% probability of acceptance (i.e., the acceptable quality level, AQL = 15%), and a 63 percent defective corresponds to a 10% probability of acceptance (i.e., the rejectable quality level, RQL = 63%). Hence, P_{95}=0.15 and P_{10} = 0.63, giving an operating ratio, OR = P_{10}/ P_{95} = 0.63/0.15 = 4.2. From Table C-1 in Appendix C of this report, a combination of f = 1/4 and i = 4 is suggested for this case. The single sampling reference plan is also obtained from Table C-1. It has an acceptance number (c) of 2. The sample size is obtained by solving the equation, $nP_{.95} = 1.263$:

$n = 1.263 / P.95 = 1.263/0.15 = 8.42 \approx 9$

Thus, the SkSP-2 plan consists of the following:

- n = 9 and c = 2 for the reference single sampling plan
- f = 1/4 and i = 4 for skip sampling

Step 2: Determine the Benefit of SkSP-2 in terms of Reduced Acceptance Sampling

Case A: Contractor Delivering High-Quality Product (having low percent defective)

Assume a contractor having a good track record. Suppose that the historical state-wide average percent defective for the contractor is five percent (i.e., 95 PWL). From Figure 5.3, the probability of accepting a lot with 5 PD using the agency's existing sampling plan is 99.5% (i.e., P = 99.5%). Using the mathematical formulas that have been discussed earlier in Section 5.3, the parameters of the equivalent SkSP-2 plan are as follows:

- The average number of lots inspected during qualification inspection,

  $U = (1 – P^i)/[P^i(1 – P)] = (1 – 0.995^4)/[0.995^4(1 – 0.995)] = 5$ lots

- The average number of lots inspected during skipping inspection,

  $V = 1/[f(1 – P)] = 1/[0.25(1 – 0.995)] = 800$ lots

- The average fraction of total lots that are inspected,

  $F = (U + fV)/(U + V) = (5 + 0.25*800)/(5 + 800) = 0.255$

- The average sample number of this SkSP-2 plan,

  $ASN(SkSP) = ASN(R)×F = 9×0.255 = 2.3$

- The probability of acceptance of SkSP-2,

  Pa (f, i) = [fP + (1 – f)*Pi]/[f + (1 – f)Pi] = [0.25*0.995 + (1-0.25)0.9954]/[0.25 + (1-0.25)*0.9954] = 0.999

- The increase in probability of accepting a lot with five percent defective,

  $P_a – P = 0.999 – 0.995 = 0.004 = 0.4\%$

Thus, for a historical quality level of five percent defective, the average fraction of total lots that are inspected is 25.5 percent, and the average sample size is 2.3 per lot. Compared to the agency's original sampling plan which has a sample size of five per lot, the use of SkSP-2, in this case, saves 54 percent of the agency's sampling and testing workload (sample size of 5 vs. average sample size of 2.3).

Case B: Contractor Delivering Poor-Quality Product (having high percent defective)

Assume a contractor having a poor track record. Suppose that the historical state-wide average percent defective for the contractor is forty percent (i.e., 60 PWL). From

Figure 5.3, the probability of accepting a lot with 40 PD using the agency's existing sampling plan is 49% (i.e., P = 0.49). Using the mathematical formulas that have been discussed earlier in Section 5.3, the parameters of the equivalent SkSP-2 plan are as follows:

- The average number of lots inspected during qualification inspection,

  $U = (1 – P^i)/[P^i(1 – P)] = (1 – 0.49^4)/[0.49^4(1 – 0.49)] = 32$ lots

- The average number of lots inspected during skipping inspection,

  V = 1/[f(1 – P)] = 1/[0.25(1 – 0.49)] = 8 lots

- The average fraction of total lots that are inspected,

  F = (U + fV)/(U + V) = (5 + 0.25*800)/(5 + 800) = 0.85

- The average sample number of this SkSP-2 plan,

  ASN(SkSP) = ASN(R)×F = 9×0.255 = 7.65

- The probability of acceptance of SkSP-2,

  Pa (f, i) = [fP + (1 – f)*Pi]/[f + (1 – f)Pi] = [0.25*0.995 + (1-0.25)0.9954]/[0.25 + (1-0.25)*0.9954] = 0.5652

- The increase in probability of accepting a lot with five percent defective,

  $P_a – P = 0.5652 – 0.49 = 0.0752\%$

Thus, for a historical quality level of forty percent defective, the average fraction of total lots that are inspected is 85 percent, and the average sample size is 7.65 per lot. Compared to the agency's original sampling plan which has a sample size of five per lot, the use of SkSP-2, in this case, will increase the agency's sampling and testing workload by 53 percent (average sample size of 5 vs. average sample size of 7.65).

The above parameters were computed for various historical quality levels (expressed in terms of percent defective) and the results are summarized in table 5.1. The last column in this table represents the percent reduction or increase in the agency's acceptance sampling and testing workload. Ratio of Skip-Lot sampling size to original sample size is found out. It can be seen that for a historical percent defective of 30 or more, the ratio is more than 1 (which represent an increase in sampling and testing requirements).

To demonstrate the effect of this SkSP-2 plan on amount of acceptance testing, the average sample number and the fraction of lots subject to inspection were plotted against various levels of historical quality. Figure 5.4 shows that when the quality level is worse than some threshold value, the average sample size of the SkSP-2 plan will exceed the sample size specified in the original acceptance plan (i.e., n = 5). In this example, this threshold is 28 percent defective. Therefore, in this example, for SkSP-2 to be effective in reducing the amount of acceptance sampling and testing, the historically achieved percent defective should be less than 29 percent (i.e., PWL greater than 71 percent).

Figure 5.5 illustrates the fraction of lots subject to inspection with the variation of historical quality levels. When historically quality levels are poor, the SkSP-2 scheme becomes very close to regular inspection (i.e., requiring inspection of every lot).

Table 5.1: Summary of SkSP-2 Calculations for the Example Problem

Percent Defective (%)	P	i	f	U	V	F	Pa	ASN (Skip-Lot)	ASN (Original)	Skip-Lot ASN: Original ASN
1	0.999	4	0.25	4	4000	0.251	0.99975	2.26	5.00	0.45
5	0.995	4	0.25	4	800	0.254	0.99873	2.28	5.00	0.46
10	0.99	4	0.25	4	400	0.257	0.99742	2.32	5.00	0.46
20	0.9	4	0.25	5	40	0.333	0.96631	3.00	5.00	0.60
30	0.71	4	0.25	10	14	0.563	0.83545	5.06	5.00	1.01
40	0.49	4	0.25	32	8	0.850	0.56520	7.65	5.00	1.53
50	0.28	4	0.25	225	6	0.981	0.29304	8.82	5.00	1.76
60	0.13	4	0.25	4023	5	0.999	0.13074	8.99	5.00	1.80
70	0.04	4	0.25	406900	4	1.000	0.04001	9.00	5.00	1.80

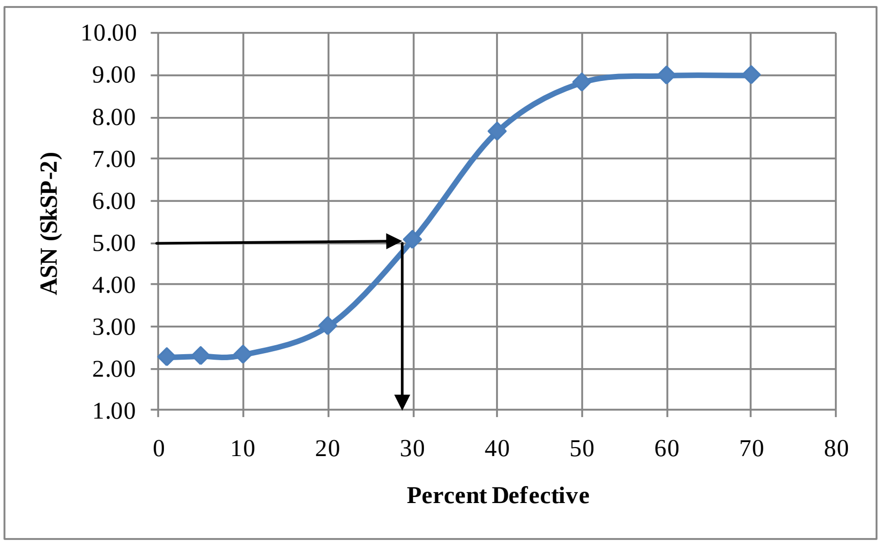

Figure 5.4: Average Sampling Numbers (i.e. Sample Size) for Different Historical Quality Levels

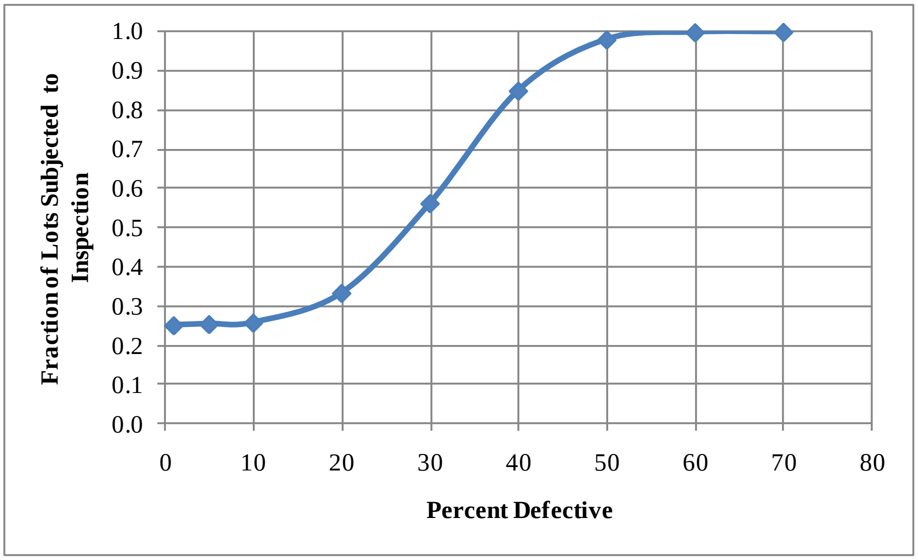

Figure 5.5: Fraction of Lots Subject to Acceptance Inspection for Different Historical Quality Levels

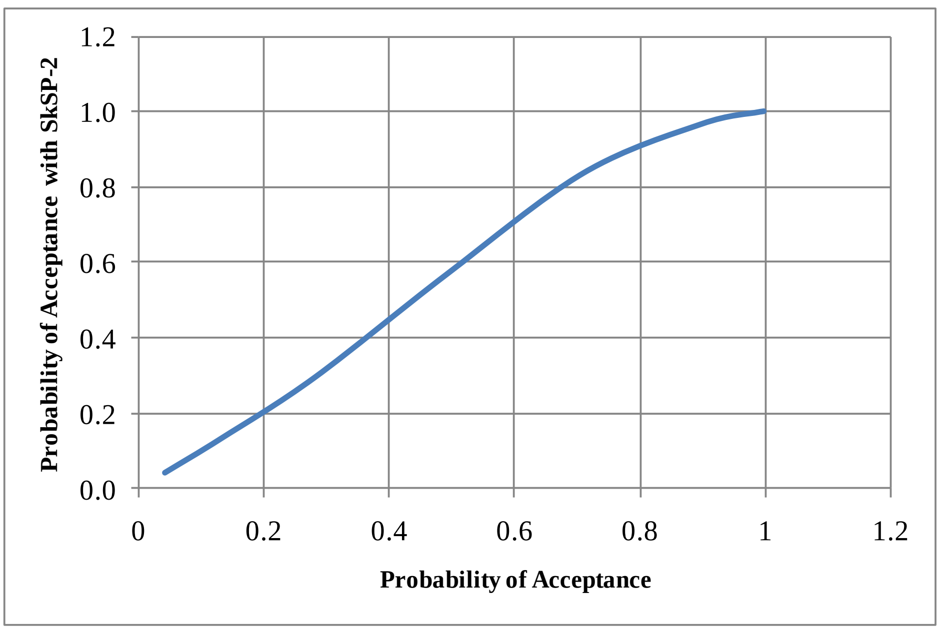

Figure 5.6: Probability of Acceptance (Reference Plan) vs Probability of Acceptance (SkSP-2)

6. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

6.1. Summary

Currently several state Departments of Transportation (DOTs) are using contractor test results, in conjunction with verification test results, for construction and materials acceptance purposes. The use of contractor test results in acceptance decisions, was codified in 1995 in Title 23, Part 637, Code of Federal Regulations (23 CFR Part 637), the Federal Highway Administration's (FHWA's) Quality Assurance Procedures for Construction *(1)*.

Making use of contractor's test results (ideally used for process control) in acceptance and payment decisions have somewhat helped state DOTs to deal with shrinking workforce and lessen the workload involved in acceptance sampling and testing. But, the practice itself has been controversial. Research has shown that the issue of bias in contractor test results is a concern *(2)*. Data gathered as part of the National Cooperative Highway Research Program (NCHRP) study 10-58(02) have shown the need for improvements and alternatives to this practice *(2)*.

This research (as documented in this thesis) reveals the conceptual and technical pitfalls of using contractor test results for acceptance and payment purposes; identifies and ranks potential alternatives and improvements to the use of contractor test results for acceptance and payment purposes; and investigates the potential application of skip-lot sampling as a means for reducing acceptance sampling and testing for highway agencies.

6.2. Conclusions

Based on the results of this study, it can be concluded that while the reasons for using contractor test results for construction and materials acceptance purposes are real (essentially shortage of state DOT staff and intensive construction schedules), the practice itself has fundamental pitfalls.

Specific conclusions regarding the verification processes of contractor's test results are:

- The statistical verification tests (F and t tests) fail to detect potential manipulations in the contractor's test results; which can lead to accepting poor-quality material and assigning unwarranted pay increase.
- For several simulated cases of undetected manipulations of contractor test results, the unwarranted increase in the pay factor ranged between 2% and 12%.

Several alternatives and improvements have been identified and evaluated based on feedback from an expert panel. Most promising alternatives were then determined based on the panel evaluation. These alternatives and improvements are discussed in the recommendations section of this chapter. Additionally, skip-lot sampling has been identified as an effective means for reducing sampling and testing workload for highway agencies. Key conclusions from the analysis of skip-lot sampling are:

- When the contractor has a history of high-quality materials and construction, skip-lot sampling significantly reduces sampling and testing workload as compared to lot-by-lot sampling. In one case study (discussed in section 5), skip-lot sampling reduced the agencies acceptance sampling effort by 54%.

- Skip-lot sampling is suitable for contractors with a good track record of providing high-quality product.
- Skip-lot sampling increases the probability of acceptance by a small fraction (e.g., 1%). The increase in probability of acceptance signifies a slight increase in the agency's risk of accepting nonconforming products.

6.3. Recommendations

Based on the results of this research, it is recommended that the practice of using contractor's test results for acceptance and payment purposes be improved or replaced with other alternatives. The following alternatives and improvements are recommended for further investigation and field trials:

- Alternatives and Improvements Related to Sampling Plan
    - Use larger lot sizes
    - Reduce sample size to 3 per lot.
    - Use larger lots to compare contractor vs. agency test results; F and t tests would have larger n and thus be more discerning (conditioned on normality of data).
    - Randomize the AQCs to be tested at any one location (i.e., do not test all AQCs at all locations).
- Alternatives and Improvements Related to Testing Administration
    - Use third–party testing for acceptance (e.g. by commercial lab representing the agency)

  - Use automated equipment and plant records to replace/decrease testing of asphalt content, gradation, air content, strength, etc.
  - Test contractors with "A" ratings at a lower frequency than contractors with "C" ratings. Contractor ratings are for quality management purposes only, with no effect on bidding.
  - Require certain certification and/or training of the contractor's technicians who perform acceptance testing.
  - Require contractor's testing staff to report to a separate unit within the contractor's organization (i.e., require a separation between the contractor's quality management team and project management team).
- Alternatives and Improvements Related to Contracting
  - Use warranties

Finally, skip-lot sampling is also a promising alternative to the use of contractor test results for acceptance purposes and is recommended for further investigation and field trials.

REFERENCES

1. Code of Federal Regulations. Quality Assurance Procedures for Construction. Final Rule, 23 CFR 637, *Federal Register*, Vol. 60, No. 125, June 29, 1995, pp. 33712.

2. Parker, F. and Turochy, R.E. *Using the Results of Contractor-Performed Tests in Quality Assurance*. Contractor's Final Report for NCHRP Project 10-58(2), Web-only Document 115, Transportation Research Board, Washington, D.C., 2006.

3. *Implementation Manual for Quality Assurance*. Quality Construction Task Force, American Association of State Highway and Transportation Officials (AASHTO), Washington, D.C, 1996.

4. National Cooperative Highway Research Program (NCHRP) Synthesis 346, State Construction Quality Assurance Programs. Transportation Research Board, Washington, D.C., 2005.

5. LaVassar, C.J., Mahoney, J.P., Willoughby, K.A. *Statistical Assessment of Quality Assurance-Quality Control Data for Hot Mix Asphalt*. Report No. WA-RD 686.1, Washington State Department of Transportation (WSDOT), Olympia, Washington, 2009.

6. Killingsworth, B.M. and Hughes, C.S. Issues Related to Use of Contractor Quality Control Data in Acceptance Decision and Payment. *Transportation Research Record: Journal of Transportation Research Board*, Vol. 1813, 2002, pp. 249-252.

7. Burati, J.L., Weed, R.M., Hughes, C.S., Hill, H.S. Evaluation of Procedures for Quality Assurance Specifications. *Federal Highway Administration*, Publication No. FHWA-HRT-04-046, McLean, VA., 2004.

8. Davidson, R. and MacKinnon, J. Estimation and Inference in Econometrics. *Oxford University Press*, Oxford, 1993.

9. Transportation Research Board (TRB), *Glossary of Highway Quality Assurance Terms*, Washington D.C, 2009.

10. Willis, J. R. A Statistical Analysis of Georgia's HMA Quality Assurance Process. MS thesis, Auburn, Alabama, 2005.

11. Hancher, Donn E., Yuhong Wang, Kamyar C. Mahboub, *Contractor Performed Quality Control on KyTC Projects*. Kentucky Transportation Center, Univ. of Kentucky, Lexington, Kentucky, 2002.

12. Scherkenbach, W.W. The Deming Route to Quality and Productivity: Road Maps and Roadblocks. *Mercury Press,* Rockville, MD, 1986.

13. DeVor, R. E., Chang, T.H., and Sutherland, J.W. Statistical Quality Design and Control, *Macmillan*, New York, 1992.

14. Gharaibeh, N.G., Garber, S.I., and Liu, L. Determining Optimum Sample Size for Percent-within-Limits Specifications. *Transportation Research Record, Journal of the Transportation Research Board (in review).*

15. Dodge, H.F. Skip-lot Sampling Plan. *Industrial Quality Control*, Vol. 11, No. 5, 1955, pp.3-5.

16. Stephens, K. S. How to Perform Skip-Lot and Chain Sampling. *American Society for Quality Control, Quality Press*, Milwaukee, Wisconsin, 1995.

17. Dodge H.F, Perry R.L. A System of Skip-Lot Plans for Lot-by-Lot Inspection. *American Society for Quality Control Technical Conference Transactions*, Milwaukee, Wisconsin, 1971, pp. 469-477.

18. Perry, R.L. Skip-Lot Sampling Plans, *Journal of Quality Technology*. Vol. 5, No. 3, July 1973, pp. 123-130.

19. Parker, R.D. and Kessler, L. A Modified Skip-Lot Sampling Plan. *Journal of Quality Technology*, Vol. 13, No. 1, 1981, pp.31-35.

20. Anthony, R.M. Analyzing Sampling Methodologies in Semiconductor Manufacturing. M.S. Thesis, Massachusetts Institute of Technology, 2004.

21. Perry R. L. A system of Skip-Lot Sampling Plans for Lot Inspection, PhD Thesis, Rutgas-The. State University, New Jersey, 1970.

22. American Association of State Highway and Transpsrtation Officials (AASHTO) *Standard Recommended Practice for Acceptance Sampling Plans for Highway Construction*, Washington D.C, 1990.

APPENDIX A: SIMULATED STATISTICAL DATA

Table A-1: Contractor's Asphalt Content Data for Sample Size Ration of 1:1 (Reducing SD)

Test No	Original Test Results	Manipulated Data				
	SD = 0.22%	Case1: SD =0.18%	Case2: SD=0.14%	Case3: SD =0.10%	Case4: SD= 0.08%	Case5: SD= 0.06%
1	4.98	4.88	5.09	5.07	4.89	4.82
2	4.73	5.00	4.97	4.92	5.01	4.96
3	5.22	5.17	4.78	4.81	4.82	4.98
4	4.69	4.79	4.94	4.81	4.95	4.91
5	4.93	4.71	4.75	4.93	4.85	4.87
Sample Mean	4.910	4.909	4.906	4.909	4.905	4.908
Sample Std. Dev.	0.212	0.184	0.141	0.109	0.076	0.065
p-value for F-test	0.966	0.758	0.423	0.211	0.067	0.039
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**	**Yes**
p-value for t-test	0.967	0.972	0.985	0.964	0.991	0.969
Reject H0 for t-test?	**No**	**No**	**No**	**No**	**No**	**No**
Quality Index (L)	0.430	0.455	0.494	0.535	0.552	NA
Quality Index (U)	3.526	3.762	4.146	4.408	4.666	NA
PWL(L)	66	67	68.33	70	70.33	NA
PWL(U)	100	100	100	100	100	NA
PWL	66	67	68.33	70	70.33	NA
Pay Factor	**88**	**88.5**	**89.165**	**90**	**90.165**	**NA**

Table A-2: Contractor's Asphalt Content Data for Sample Size Ration of 1:1 (Increasing Mean)

Test No	Original Test Results	Manipulated Data				
	Mean = 4.9%	Case1: Mean =5%	Case2: Mean = 5.13%	Case3: Mean = 5.18%	Case4: Mean= 5.21%	Case5: Mean= 5.21%
1	5.03	4.79	5.14	4.93	5.36	5.55
2	4.80	4.91	5.00	5.25	5.17	5.24
3	4.60	5.02	5.41	5.13	5.09	5.06
4	4.88	5.38	4.84	5.53	5.51	5.36
5	5.20	5.09	5.25	5.06	4.93	4.98
Sample Mean	4.904	5.038	5.130	5.181	5.210	5.238
Sample Std. Dev.	0.226	0.224	0.223	0.225	0.226	0.227
p-value for F-test	0.956	0.970	0.977	0.961	0.954	0.945
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**	**No**
p-value for t-test	0.995	0.365	0.144	0.084	0.061	0.046
Reject H0 for t-test?	**No**	**No**	**No**	**No**	**No**	**No**
Quality Index (L)	0.400	0.683	0.819	0.870	0.894	NA
Quality Index (U)	3.412	2.944	2.513	2.263	2.125	NA
PWL(L)	65	71.33	79	80.5	81.33	NA
PWL(U)	100	100	99.8	99.4	99.1	NA
PWL	65	71.33	78.8	79.9	80.43	NA
Pay Factor	**88**	**88.5**	**89.165**	**94.95**	**95.215**	**NA**

Table A-3: Contractor's Asphalt Content Data for Sample Size Ration of 1:2 (Reducing SD)

Test No.	Original Test Results	Manipulated Test Results				
	SD=0.213%	Case 1 SD= 0.207%	Case 2 SD =0.177%	Case 3 SD=0.149%	Case 4 SD =0.105%	Case 5 SD= 0.098%
1	5.02	5.06	5.03	4.97	4.97	4.94
2	5.18	5.07	5.08	5.07	5.02	5.00
3	4.52	4.62	4.66	4.73	4.74	4.76
4	4.88	4.90	4.87	4.88	4.89	4.88
5	4.95	4.94	4.91	4.93	4.92	4.92
6	4.97	4.98	4.96	4.97	4.98	4.97
7	5.22	5.32	5.24	5.18	5.07	5.07
8	4.71	4.67	4.72	4.71	4.79	4.79
9	4.80	4.84	4.82	4.83	4.86	4.86
10	4.78	4.78	4.77	4.79	4.83	4.82
Sample Mean	4.902	4.918	4.906	4.906	4.907	4.901
Sample Std. Dev.	0.213	0.207	0.177	0.149	0.105	0.098
p-value for F-test	0.877	0.826	0.562	0.322	0.065	0.044
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**	**Yes**
p-value for t-test	0.985	0.905	0.985	0.984	0.971	0.970
Reject H0 for t-test?	**No**	**No**	**No**	**No**	**No**	**No**
Quality Index (L)	0.400	0.461	0.465	0.512	0.559	NA
Quality Index (U)	3.479	3.490	3.898	4.291	5.028	NA
PWL(L)	65.3	67.3	67.9	69.3	71	NA
PWL(U)	100	100	100	100	100	NA
PWL	65.3	67.3	67.9	69.3	71	NA
Pay Factor	**87.65**	**88.65**	**88.95**	**89.65**	**90.5**	**NA**

Table A-4: Contractor's Asphalt Content Data for Sample Size Ration of 1:2 (Increasing Mean)

Test No.	Original Test Results	Manipulated Data				
	Mean=4.902%	Case 1 Mean = 5.085%	Case 2 Mean = 5.115%	Case 3 Mean = 5.130%	Case 4 Mean = 5.149%	Case 5 Mean= 5.169%
1	5.02	4.99	5.04	5.04	5.03	4.88
2	5.18	5.25	5.20	5.12	4.94	5.00
3	4.52	4.96	5.08	4.97	5.21	5.30
4	4.88	4.90	5.29	5.33	5.30	4.91
5	4.95	5.21	4.83	5.15	4.99	5.09
6	4.97	5.05	4.95	5.54	5.38	5.18
7	5.22	5.33	5.49	5.25	5.56	5.14
8	4.71	5.39	5.16	5.19	5.12	5.26
9	4.80	4.67	4.76	4.90	5.17	5.37
10	4.78	5.11	5.34	4.80	4.79	5.57
Sample Mean	4.902	5.085	5.115	5.130	5.149	5.169
Sample Std. Dev.	0.213	0.218	0.229	0.217	0.228	0.214
p-value for F-test	0.857	0.903	0.985	0.890	0.984	0.866
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**	**No**
p-value for t-test	0.988	0.153	0.111	0.080	0.068	0.042
Reject H_0 for t-test?	**No**	**No**	**No**	**No**	**No**	**Yes**
Quality Index (L)	0.420	0.896	0.932	0.989	0.996	NA
Quality Index (U)	3.496	2.605	2.391	2.389	2.227	NA
PWL(L)	65.3	81.5	82.25	83.75	84	NA
PWL(U)	100	99.5	99.3	99.3	99.1	NA
PWL	65.3	81	81.55	83.05	83.1	NA
Pay Factor	**87.65**	**95.5**	**95.775**	**96.525**	**96.55**	NA

Table A-5: Contractor's Asphalt Content Data for Sample Size Ration of 1:4 (Reducing SD)

Test No	Original Test Results	Manipulated Data				
	SD = 0.22%	Case1: SD = 0.2%	Case2: SD= 0.17%	Case3: SD= 0.13%	Case4: SD= 0.12%	Case5: SD=0.11%
1	5.19	4.81	4.91	4.80	4.81	5.10
2	4.78	5.07	4.79	4.67	4.96	5.06
3	4.98	5.06	5.12	4.77	4.93	4.95
4	4.66	5.13	4.77	4.74	4.89	4.81
5	4.75	4.60	5.00	4.99	5.05	4.77
6	4.49	4.85	5.05	4.86	4.96	4.66
7	5.01	4.50	4.75	5.07	4.77	5.04
8	4.59	4.95	4.85	5.11	4.87	4.82
9	4.84	5.01	4.62	4.87	4.91	4.94
10	4.78	4.67	4.97	5.15	4.75	4.97
11	4.85	4.92	4.87	5.03	4.98	5.00
12	5.17	4.88	4.71	4.81	4.81	4.84
13	5.37	4.78	5.04	4.84	4.93	4.75
14	5.11	4.96	4.83	4.92	5.01	4.99
15	4.69	5.16	5.27	4.94	5.02	4.85
16	4.94	4.75	5.11	4.91	5.17	4.93
17	4.90	5.36	4.95	4.69	4.65	4.86
18	5.04	4.99	4.93	4.96	4.85	4.87
19	4.90	4.71	4.66	4.89	4.83	4.89
20	5.08	4.86	4.89	4.98	4.87	4.91
Sample Mean	4.940	4.900	4.900	4.900	4.900	4.900
Sample Std. Dev.	0.220	0.206	0.166	0.133	0.119	0.111
p-value for F-test	0.843	0.716	0.345	0.114	0.053	0.031
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**	**Yes**
p-value for t-test	0.974	0.988	0.985	0.978	0.977	0.977
Reject H0 for t-test?	**No**	**No**	**No**	**No**	**No**	**No**
Quality Index (L)	0.400	0.399	0.486	0.544	0.584	NA
Quality Index (U)	3.479	3.509	4.125	4.809	5.157	NA
PWL(L)	65.3	65.300	68.900	70.300	71.900	NA
PWL(U)	100	100.000	100.000	100.000	100.000	NA
PWL	65.3	65.300	68.900	70.300	71.900	NA
Pay Factor	**87.65**	**87.65**	**89.45**	**90.15**	**90.835**	**NA**

Table A-6: Contractor's Asphalt Content Data for Sample Size Ration of 1:4 (Increasing Mean)

Test No.	Original Test Results	Manipulated Data			
	Mean = 4.9%	Case1: Mean =5%	Case2: Mean = 5.1%	Case3: Mean = 5.12%	Case4: Mean = 5.13%
1	4.99	5.25	5.39	5.18	5.22
2	4.78	4.70	5.01	5.15	4.94
3	4.65	5.29	5.38	4.95	5.19
4	5.03	4.86	5.30	5.38	5.26
5	5.11	5.22	5.47	4.99	5.03
6	5.20	4.82	5.20	4.80	4.98
7	4.74	5.16	5.24	5.13	5.15
8	5.35	4.80	4.63	5.25	4.97
9	4.55	4.96	4.97	4.90	4.82
10	4.51	5.48	4.94	5.07	5.28
11	4.95	5.07	5.15	5.27	4.71
12	4.71	5.14	5.17	5.22	5.53
13	4.81	5.00	5.10	5.35	5.10
14	4.84	4.74	5.12	4.85	5.07
15	4.87	4.92	5.02	5.56	5.16
16	4.90	4.90	5.06	5.11	5.36
17	4.91	5.11	4.89	4.74	5.10
18	4.97	4.61	4.85	5.03	5.37
19	5.06	5.02	4.76	5.06	4.90
20	5.16	5.05	5.25	5.43	5.48
Sample Mean	4.904	5.005	5.095	5.120	5.132
Sample Std. Dev.	0.217	0.220	0.219	0.218	0.214
p-value for F-test	0.833	0.862	0.852	0.841	0.811
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**
p-value for t-test	0.968	0.388	0.102	0.064	0.049
Reject H0 for t-test?	**No**	**No**	**No**	**No**	**Yes**
Quality Index (L)	0.398	0.757	1.044	1.120	NA
Quality Index (U)	3.361	2.897	2.471	2.353	NA
PWL(L)	65.300	77.300	85.250	87.600	NA
PWL(U)	100.000	99.700	99.300	99.200	NA
PWL	65.300	77.000	84.550	86.800	NA
Pay Factor	87.65	93.5	97.275	98.4	**NA**

Table A-7: Contractor's Asphalt Content Data for Sample Size Ration of 1:10 (Reducing SD)

Test No.	Original Test Results	Manipulated Data				
	SD = 0.22%	Case1: SD= 0.20%	Case2: SD= 0.17%	Case3: SD= 0.14%	Case4: SD= 0.13%	Case5: SD= 0.12%
1	5.15	5.18	4.90	4.94	4.78	5.04
2	5.05	5.33	4.88	4.98	5.04	4.82
3	5.21	4.95	4.80	4.97	4.79	4.97
4	5.26	4.91	4.90	4.60	4.97	4.81
5	4.83	5.02	4.85	4.67	4.80	4.69
6	4.89	4.92	4.92	4.77	4.91	4.73
7	4.86	4.81	4.87	4.75	4.94	4.77
8	4.98	4.95	4.86	4.74	4.90	4.84
9	4.88	4.61	4.94	5.09	4.85	4.96
10	5.35	4.68	5.01	5.01	4.75	4.83
11	4.53	5.00	4.83	4.84	4.65	4.90
12	5.36	4.86	4.61	5.27	4.92	4.81
13	4.82	4.72	4.79	4.93	4.82	5.07
14	5.10	4.88	4.75	4.93	4.93	4.85
15	4.64	4.57	4.88	4.92	5.29	4.66
16	4.92	4.86	4.98	5.14	4.87	4.96
17	5.24	5.18	5.18	5.15	5.00	5.03
18	4.88	4.79	5.02	4.82	4.95	4.87
19	4.78	5.12	5.06	4.85	4.92	4.77
20	5.07	4.90	4.97	4.80	4.88	4.91
21	4.76	5.04	4.77	4.97	5.09	4.70
22	5.07	4.76	5.15	4.79	4.89	4.98
23	4.79	4.98	4.66	4.80	4.86	5.06
24	4.71	4.68	4.60	4.73	4.76	4.75
25	4.60	4.66	5.05	5.11	4.84	4.88
26	4.94	4.55	4.70	4.99	4.98	5.00
27	4.85	4.88	5.00	4.86	5.02	4.90
28	4.48	4.77	4.67	5.06	4.96	5.00
29	4.74	4.62	5.07	4.96	5.01	4.98
30	4.72	4.90	4.78	5.03	4.88	4.94
31	4.80	5.11	4.50	4.90	4.83	4.63
32	4.99	5.10	5.00	5.05	5.04	4.92
33	4.96	4.80	4.95	4.85	4.84	5.01
34	5.02	4.75	4.93	4.69	5.16	4.86
35	4.94	4.97	4.98	4.62	5.02	4.80
36	4.58	5.30	4.93	4.87	4.74	4.79
37	4.66	5.14	5.04	4.88	4.81	5.12
38	5.11	5.02	4.80	4.95	4.70	4.82
39	4.75	4.99	5.09	4.78	4.95	4.89
40	4.93	4.73	5.13	5.06	4.98	5.14
41	4.90	4.46	4.71	4.83	5.05	5.10
42	4.66	5.06	5.24	5.00	5.11	4.86
43	4.45	4.93	4.84	4.91	4.86	5.03
44	5.00	5.08	4.74	4.91	5.07	4.94
45	5.16	5.03	5.38	4.95	4.61	5.17
46	5.12	4.73	4.73	5.02	4.72	4.88
47	4.84	4.82	4.83	4.72	4.67	4.93
48	4.69	4.83	4.82	4.88	4.96	4.74
49	5.01	4.85	4.95	4.83	4.81	4.95
50	5.03	5.21	5.11	4.89	4.89	4.92
Sample Mean	4.901	4.900	4.903	4.901	4.902	4.900
Sample Std. Dev.	0.217	0.196	0.174	0.141	0.135	0.127
p-value for F-test	0.784	0.577	0.363	0.109	0.079	0.048
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**	**Yes**
p-value for t-test	0.984	0.978	0.997	0.978	0.989	0.963
Reject H0 for t-test?	**No**	**No**	**No**	**No**	**No**	**No**
Quality Index (L)	0.376	0.411	0.468	0.552	0.579	NA
Quality Index (U)	3.336	3.664	4.063	4.886	5.066	NA
PWL(L)	65.3	66.000	68.000	71.000	71.900	NA
PWL(U)	99.8	100.000	100.000	100.000	100.000	NA
PWL	65.1	66.000	68.000	71.000	71.900	NA
Pay Factor	**87.55**	**88**	**89**	**90.5**	**90.95**	**NA**

Table A-8: Contractor's Asphalt Content Data for Sample Size Ration of 1:10 (Increasing Mean)

Test No.	Original Test Results	Manipulated Data			
	Mean = 4.9	Case1: Mean =5	Case2: Mean =5.1	Case3: Mean= 5.115	Case4: Mean =5.12
1	4.87	4.93	4.81	5.21	5.25
2	4.82	5.11	5.22	5.37	5.20
3	4.93	5.38	5.05	4.80	5.30
4	5.04	4.70	5.40	5.20	5.14
5	5.08	4.74	5.13	4.99	5.18
6	4.63	5.31	5.33	4.76	5.03
7	4.78	5.03	5.66	5.39	5.12
8	4.93	5.13	5.00	5.13	5.26
9	4.97	5.12	4.94	4.93	5.31
10	5.20	4.88	4.95	5.18	5.39
11	4.92	4.90	5.02	5.51	5.23
12	5.13	5.08	4.90	5.34	4.87
13	4.99	4.66	4.66	5.18	5.19
14	5.00	5.06	5.18	4.83	5.05
15	4.57	5.28	5.20	5.28	5.33
16	4.79	4.56	4.76	5.22	5.11
17	4.80	4.51	4.87	4.69	5.28
18	5.26	5.27	5.13	4.98	5.08
19	4.71	4.78	4.98	5.02	5.17
20	4.90	4.85	5.35	5.26	5.06
21	5.01	5.26	5.37	5.10	5.01
22	4.37	5.05	5.46	5.07	5.07
23	5.10	4.94	5.29	5.71	4.65
24	5.23	5.23	5.16	5.43	4.97
25	4.69	4.89	4.56	5.33	4.95
26	4.66	5.45	4.75	4.92	5.16
27	5.17	4.86	5.53	4.91	5.03
28	4.84	4.84	4.92	4.63	5.22
29	4.48	4.80	5.08	5.24	5.15
30	4.82	5.19	5.11	5.25	5.11
31	4.85	5.00	5.04	5.06	5.42
32	4.74	4.63	4.99	5.47	5.09
33	4.72	5.17	5.19	5.29	5.38
34	4.88	5.08	4.90	4.95	4.99
35	5.04	4.75	5.42	5.11	5.51
36	5.12	4.99	5.22	5.17	4.81
37	4.86	5.15	4.83	5.06	4.89
38	4.89	5.48	5.30	5.15	4.92
39	5.29	5.01	5.28	4.89	5.65
40	4.77	4.77	5.06	5.30	5.35
41	5.03	4.97	5.11	5.09	4.70
42	4.67	4.97	4.85	4.97	4.86
43	4.96	4.95	5.24	5.04	4.95
44	5.15	5.17	5.09	5.40	4.82
45	4.55	5.21	5.03	4.85	5.00
46	5.07	4.91	4.96	5.15	5.43
47	4.95	5.10	5.15	5.04	5.49
48	4.60	5.02	5.08	5.01	5.27
49	4.75	5.04	5.27	5.12	4.91
50	5.55	4.82	5.25	4.86	4.77
Sample Mean	4.903	4.999	5.101	5.117	5.122
Sample Std. Dev.	0.227	0.220	0.225	0.221	0.217
p-value for F-test	0.886	0.818	0.865	0.826	0.789
Reject H_0 for F-test?	**No**	**No**	**No**	**No**	**No**
p-value for t-test	0.925	0.405	0.080	0.054	0.046
Reject H0 for t-test?	**No**	**No**	**No**	**No**	**Yes**
Quality Index (L)	0.372	0.781	1.150	1.226	NA
Quality Index (U)	3.186	2.861	2.338	2.302	NA
PWL(L)	65.300	78.250	87.600	89.000	NA
PWL(U)	99.700	99.500	99.100	99.050	NA
PWL	65.000	77.750	86.700	88.050	NA
Pay Factor	87.5	93.875	98.35	99.025	**NA**

Table A-9: Contractor's PCC Thickness Data for Sample Size Ration of 1:1 (Reducing SD)

Test No	Original Test Results	Manipulated Data					
	SD = 0.4in	Case1: SD =0.35in	Case2: SD =0.3in	Case3: SD =0.2in	Case4: SD =0.17in	Case5: SD =0.10in	Case6: SD =0.09in
1	11.73	11.32	11.29	11.00	11.15	11.24	11.24
2	11.30	10.86	11.65	11.49	11.30	11.27	11.27
3	11.21	11.70	10.99	11.21	11.11	11.37	11.36
4	10.76	11.11	11.09	11.34	11.46	11.12	11.13
Sample Mean	11.250	11.250	11.255	11.260	11.253	11.250	11.250
Sample Std. Dev	0.400	0.355	0.290	0.205	0.158	0.102	0.093
p-value for F-test	0.981	0.868	0.627	0.312	0.168	0.052	0.041
Reject H_0 for F-test?	No	No	No	No	No	No	Yes
p-value for t-test	0.971	0.969	0.982	0.998	0.972	0.960	0.960
Reject H_0 for t-test?	No	No	No	No	No	No	No
Quality Index	0.695	0.735	0.805	0.895	0.924	0.958	NA
PWL	75	76	78.33	81.25	81.75	83	NA
Pay Factor	**92.5**	**93**	**94.165**	**95.625**	**95.875**	**96.5**	NA

Table A-10: Contractor's PCC Thickness Data for Sample Size Ration of 1:1 (Increasing Mean)

Test No	Original Test Results	Manipulated Data						
	Mean = 11.25in	Case1: Mean =11.35in	Case2: Mean =11.45in	Case3: Mean =11.55in	Case4: Mean =11.65in	Case 5: Mean =11.75in	Case6: Mean =11.90in	Case7: Mean =11.95in
1	11.74	10.86	11.70	11.13	11.64	11.72	11.84	11.88
2	11.30	11.31	11.79	12.00	11.30	11.42	11.60	11.65
3	11.21	11.40	11.35	11.73	11.44	11.54	11.70	11.74
4	10.77	11.86	10.93	11.31	12.19	12.31	12.49	12.52
Sample Mean	11.255	11.357	11.442	11.542	11.644	11.747	11.907	11.947
Sample Std. Dev	0.399	0.409	0.389	0.394	0.393	0.398	0.399	0.394
p-value for F-test	0.984	0.952	0.982	0.998	0.996	0.987	0.984	1.000
Reject H_0 for F-test?	No	No	No	No	No	No	No	No
p-value for t-test	0.989	0.723	0.515	0.336	0.207	0.127	0.058	0.047
Reject H_0 for t-test?	No	No	No	No	No	No	No	Yes
Quality Index	0.689	0.808	0.921	1.000	1.064	1.103	1.141	NA
PWL	74.75	78.67	81.75	84	85.5	86.5	87.5	NA
Pay Factor	**92.375**	**94.335**	**95.875**	**97**	**97.75**	**98.25**	**98.75**	NA

Table A-11: Contractor's PCC Thickness Data for Sample Size Ration of 1:2 (Reducing SD)

Test No.	Original Test Results	Manipulated Data				
	SD=0.405 in	Case 1 SD= 0.352 in	Case 2 SD =0.239 in	Case 3 SD= 0.206 in	Case 4 SD = 0.177 in	Case 5 SD =0.157 in
1	11.27	10.64	11.33	11.07	11.22	11.18
2	10.68	11.49	11.41	10.98	11.15	11.32
3	11.49	11.81	11.20	11.41	11.35	11.39
4	10.81	11.22	11.60	11.28	11.30	11.02
5	11.07	11.46	11.43	11.17	11.37	11.20
6	11.18	11.13	11.14	11.33	10.97	11.29
7	11.91	10.99	11.00	11.19	11.52	11.52
8	11.58	11.31	10.87	11.64	11.09	11.13
Sample Mean	11.250	11.259	11.246	11.258	11.245	11.257
Sample Std. Dev	0.405	0.352	0.239	0.206	0.177	0.157
p-value for F-test	0.868	0.655	0.214	0.119	0.060	0.034
Reject H_0 for F-test?	No	No	No	No	No	Yes
p-value for t-test	0.972	1.000	0.947	0.996	0.937	0.992
Reject H_0 for t-test?	No	No	No	No	No	No
Quality Index	0.651	0.732	0.871	0.954	0.971	NA
PWL	73.75	76.3	80.5	82.75	83.25	NA
Pay Factor	**91.875**	**93.15**	**95.25**	**96.375**	**96.625**	NA

Table A-12: Contractor's PCC Thickness Data for Sample Size Ration of 1:2 (Increasing Mean)

Test No.	Original Test Results	Manipulated Data				
	Mean=11.262 in	Case 1 Mean=11.336 in	Case 2 Mean=11.560 in	Case 3 Mean=11.646 in	Case 4 Mean=11.795 in	Case 5 Mean=11.836 in
1	11.27	11.44	11.66	11.72	11.88	11.89
2	10.68	11.33	11.59	11.56	11.76	11.78
3	11.49	11.59	11.42	11.81	12.01	12.00
4	10.81	11.87	12.19	12.21	12.43	12.52
5	11.07	11.13	11.71	11.48	11.64	11.57
6	11.18	11.69	11.74	12.07	12.11	12.19
7	11.91	11.07	11.20	11.29	11.48	11.43
8	11.58	10.57	10.97	11.03	11.05	11.31
Sample Mean	11.250	11.336	11.560	11.646	11.795	11.836
Sample Std. Dev	0.405	0.412	0.370	0.391	0.420	0.403
p-value for F-test	0.868	0.893	0.732	0.816	0.925	0.862
Reject H_0 for F-test?	No	No	No	No	No	No
p-value for t-test	0.972	0.765	0.228	0.142	0.062	0.042
Reject H_0 for t-test?	No	No	No	No	No	Yes
Quality Index	0.651	0.787	1.165	1.219	1.290	NA
PWL	73.75	78.3	88	89.25	90.5	NA
Pay Factor	**91.875**	**94.15**	**99**	**99.625**	**100.25**	NA

Table A-13: Contractor's PCC Thickness Data for Sample Size Ration of 1:4 (Reducing SD)

Test No	Original Test Results	Manipulated Data				
	SD = 0.423in	Case1: SD= 0.35in	Case2: SD = 0.3in	Case3: SD = 0.25in	Case4: SD = 0.2in	Case5: SD = 0.18in
1	12.23	11.68	11.79	11.23	10.82	11.26
2	11.51	10.65	10.87	11.19	11.17	11.09
3	10.50	10.83	10.70	10.98	11.38	11.13
4	11.40	10.89	11.29	11.60	11.43	11.29
5	11.36	11.35	11.21	11.30	11.09	11.45
6	11.61	11.04	11.45	10.77	11.46	11.38
7	11.19	11.09	11.35	11.06	10.99	11.05
8	11.09	11.95	11.18	11.37	11.33	11.19
9	11.13	11.41	11.13	11.64	11.13	10.91
10	11.04	11.60	11.50	11.25	11.19	11.20
11	10.74	10.98	11.60	11.15	11.31	11.33
12	11.56	11.53	10.91	11.48	11.52	11.61
13	11.72	11.49	11.58	11.45	11.07	11.52
14	11.29	11.27	11.06	11.12	11.25	11.37
15	10.92	11.21	11.01	10.92	11.64	11.02
16	10.84	11.16	11.30	11.39	11.26	11.23
Sample Mean	11.259	11.257	11.246	11.244	11.252	11.252
Sample Std. Dev	0.423	0.345	0.296	0.243	0.209	0.190
p-value for F-test	0.886	0.547	0.336	0.141	0.062	0.033
Reject H_0 for F-test?	No	No	No	No	No	Yes
p-value for t-test	0.999	0.992	0.945	0.927	0.964	0.958
Reject H_0 for t-test?	No	No	No	No	No	No
Quality Index	0.631	0.740	0.802	0.913	1.024	NA
PWL	73.3	76.75	78.67	82.25	84.5	NA
Pay Factor	**91.65**	**93.375**	**94.335**	**96.125**	**97.25**	NA

Table A-14: Contractor's PCC Thickness Data for Sample Size Ration of 1:4 (Increasing Mean)

Test No	Original Test Results	Manipulation Data					
	Mean = 11.25in	Case1: Mean= 11.35in	Case2: Mean =11.45in	Case3: Mean =11.55in	Case4: Mean= 11.65in	Case5: Mean = 11.70in	Case6: Mean =11.74in
1	10.93	11.69	11.31	11.52	11.99	12.31	11.59
2	11.77	11.26	11.71	11.26	11.03	12.15	11.85
3	11.67	10.99	11.34	11.47	11.23	11.39	11.49
4	11.57	11.46	11.78	11.16	11.05	11.63	11.62
5	11.09	11.88	11.12	11.01	12.04	11.66	12.49
6	11.51	10.55	11.98	11.99	11.55	12.43	11.88
7	11.05	11.03	11.20	11.30	12.19	11.20	12.00
8	11.43	11.08	11.49	11.82	11.81	11.56	11.30
9	10.67	11.77	10.61	10.89	11.51	11.92	11.45
10	11.34	11.30	10.99	11.83	11.59	10.97	12.16
11	10.53	11.37	11.53	11.38	11.89	11.46	11.78
12	11.20	12.01	10.85	11.57	12.42	11.80	12.06
13	11.15	10.77	11.44	12.11	11.68	12.05	11.70
14	10.79	11.60	11.59	11.66	11.75	11.88	10.95
15	12.06	11.53	11.86	12.47	11.37	11.73	11.27
16	11.27	11.16	12.10	11.71	11.43	11.26	12.29
Sample Mean	11.252	11.340	11.431	11.572	11.657	11.712	11.742
Sample Std. Dev	0.413	0.403	0.412	0.417	0.394	0.406	0.406
p-value for F-test	0.846	0.804	0.842	0.860	0.765	0.816	0.815
Reject H_0 for F-test?	No	No	No	No	No	No	No
p-value for t-test	0.978	0.722	0.464	0.194	0.089	0.062	0.048
Reject H_0 for t-test?	No	No	No	No	No	No	Yes
Quality Index	0.630	0.820	0.973	1.199	1.376	1.419	NA
PWL	73.3	79.25	83.25	88.65	91.7	92.5	NA
Pay Factor	**91.65**	**94.625**	**96.625**	**99.325**	**100.85**	**101.25**	NA

Table A-15: Contractor's PCC Thickness Data for Sample Size Ration of 1:10 (Reducing SD)

Test No	Original Test Results	Manipulated Data				
	SD = 0.405in	Case1: SD =0.35in	Case2: SD =0.3in	Case3: SD =0.25in	Case4: SD =0.22in	Case5: SD =0.2in
1	10.81	11.36	11.66	11.26	11.07	10.80
2	11.61	11.24	11.43	11.45	11.14	11.23
3	11.82	11.16	10.57	11.36	11.35	11.16
4	10.86	11.57	11.50	11.17	11.63	11.48
5	12.30	11.18	11.16	11.60	11.34	11.50
6	11.11	11.01	11.28	11.07	11.26	11.44
7	11.57	11.52	11.36	11.02	11.03	11.22
8	11.02	10.99	11.14	11.43	11.25	10.84
9	11.32	11.65	11.50	11.34	11.00	10.98
10	11.13	11.45	11.45	11.47	11.46	11.20
11	11.98	11.08	10.82	11.49	11.27	11.34
12	10.76	10.94	10.74	11.21	11.38	11.09
13	11.89	11.10	10.91	11.31	11.54	11.38
14	11.71	11.32	11.61	11.24	11.04	11.02
15	11.49	11.58	11.99	11.01	11.42	11.06
16	11.45	11.31	11.58	11.12	11.18	11.18
17	11.41	11.19	11.45	11.41	11.47	11.19
18	11.34	10.92	10.94	11.27	11.19	11.40
19	11.54	10.83	10.97	11.23	11.31	11.35
20	11.52	10.63	11.07	11.52	11.22	11.25
21	11.25	11.39	11.30	11.18	11.29	11.04
22	11.26	11.12	11.25	11.39	11.13	10.98
23	11.16	11.06	11.22	11.08	11.32	11.12
24	10.97	10.77	11.74	10.83	11.42	11.10
25	10.61	10.97	11.39	11.20	11.40	11.53
26	11.68	11.25	11.32	11.29	10.94	11.32
27	10.90	11.22	11.33	11.66	11.08	11.46
28	10.59	11.37	10.86	11.55	10.84	11.30
29	11.39	11.50	11.82	11.10	11.23	11.56
30	11.65	11.73	11.38	11.32	11.76	10.94
31	10.92	10.87	11.01	11.37	11.11	11.24
32	11.18	11.87	11.18	10.85	11.36	11.62
33	11.36	11.41	11.03	10.92	11.51	11.73
34	10.44	11.28	10.88	11.15	11.56	11.15
35	10.71	11.96	11.07	11.76	10.97	11.38
36	11.22	10.71	11.13	10.96	11.15	11.28
37	11.05	11.48	11.25	11.05	10.91	11.31
38	11.00	10.51	11.55	10.99	11.65	11.13
39	11.07	11.76	11.21	11.72	11.19	11.42
40	11.30	11.67	11.11	10.66	10.68	11.27
Sample Mean	11.259	11.248	11.254	11.250	11.250	11.250
Sample Std. Dev	0.406	0.341	0.304	0.249	0.232	0.209
p-value for F-test	0.782	0.483	0.315	0.112	0.072	0.032
Reject H_0 for F-test?	No	No	No	No	No	No
p-value for t-test	0.999	0.955	0.977	0.950	0.950	0.940
Reject H_0 for t-test?	No	No	No	No	No	Yes
Quality Index	0.645	0.727	0.822	0.962	1.019	NA
PWL	73.67	76.67	79.33	83.3	84.6	NA
Pay Factor	**91.835**	**93.335**	**94.665**	**96.65**	**97.3**	NA

Table A-16: Contractor's PCC Thickness Data for Sample Size Ration of 1:10 (Increasing Mean)

Test No	Original Test Results	Manipulated Data				
	Mean = 11.25in	Case1: Mean =11.35in	Case2: Mean =11.45in	Case3: Mean =11.55in	Case4: Mean =11.65in	Case5: Mean =11.68in
1	11.41	11.44	11.48	11.37	12.11	12.01
2	10.98	11.42	11.26	11.69	11.92	11.70
3	11.37	11.30	11.58	11.91	11.48	11.37
4	10.61	11.49	11.14	11.79	11.25	11.54
5	10.72	11.47	10.51	11.68	12.77	11.50
6	11.43	11.02	11.38	11.58	11.74	12.20
7	10.89	11.30	11.03	11.42	11.55	11.78
8	10.48	11.77	12.06	10.89	11.89	11.90
9	11.73	11.17	10.81	11.29	11.71	11.66
10	11.28	11.75	11.57	11.74	12.03	12.15
11	11.34	11.15	11.35	11.22	11.53	11.10
12	10.86	11.01	11.82	11.62	11.66	11.87
13	11.40	11.14	11.69	11.25	11.80	11.59
14	11.09	11.87	11.75	11.56	10.71	11.41
15	11.59	11.05	12.17	11.64	11.99	12.33
16	11.64	10.90	11.99	10.89	11.62	11.49
17	11.70	10.85	11.23	11.35	11.07	11.98
18	10.79	11.83	11.84	11.50	11.52	11.83
19	11.88	10.72	10.72	11.45	11.95	11.78
20	10.97	11.63	10.91	11.53	12.31	11.65
21	12.08	11.67	11.35	12.00	12.12	11.02
22	11.27	11.21	11.19	11.85	12.25	11.62
23	10.83	11.39	11.63	11.95	10.91	12.05
24	10.17	11.35	10.99	11.20	11.68	12.07
25	11.97	10.78	11.72	12.13	11.96	11.30
26	11.02	12.08	11.43	11.71	11.45	11.82
27	11.32	11.53	11.55	11.51	11.39	10.80
28	11.11	10.60	11.29	11.43	11.33	11.13
29	11.51	11.56	11.05	12.03	11.11	12.51
30	11.48	12.23	11.64	11.34	11.16	12.41
31	11.24	11.59	11.47	11.00	11.80	11.46
32	11.05	10.52	11.87	12.09	11.58	11.19
33	10.93	11.94	12.25	12.25	11.19	11.73
34	11.60	11.37	11.94	12.45	11.76	12.12
35	11.17	11.09	11.31	11.04	11.30	11.32
36	11.77	11.60	11.50	11.78	11.41	11.42
37	11.14	11.27	11.40	10.64	11.38	11.94
38	11.52	11.23	11.75	11.15	11.65	11.71
39	11.20	11.70	11.15	11.84	12.19	11.58
40	11.19	10.97	11.10	11.13	11.84	11.26
Sample Mean	11.243	11.348	11.446	11.548	11.652	11.683
Sample Std. Dev	0.409	0.397	0.397	0.399	0.414	0.391
p-value for F-test	0.796	0.740	0.742	0.749	0.815	0.715
Reject H_0 for F-test?	No	No	No	No	No	No
p-value for t-test	0.943	0.670	0.374	0.175	0.077	0.046
Reject H_0 for t-test?	No	No	No	No	No	Yes
Quality Index	0.604	0.863	1.080	1.291	1.451	NA
PWL	72.33	80.5	86	90.25	92.9	NA
Pay Factor	**91.165**	**95.25**	**98**	**100.125**	**101.45**	NA

APPENDIX B: PANEL RESPONSES TO ALTERNATIVES

Responses from all the panel members are clubbed together.

- Approprite for states that are looking to reduce the agency's workload, but are not interested in using CAT.

No.	Alternatives to CAT	Potential for Reducing Agency's workload	Potential for Increasing Agency's Risk of Accepting Poor Quality Products	Ease of Implementation	Worthy of Further Investigation?
		(Low, Med., High)	(Low, Med.,High)	(Low, Med.,High)	(Yes, Maybe, No)
1	Start project with normal test frequency, and then increase the lot size once there is evidence that the contractor's process is under control (Florida's approach)	Low, Low-Med, Med, High, Med	Low, Low, Med, Med, Low	High, High, Low, Low, High	Yes, Maybe, Maybe, Yes, yes
2	Start project with normal test frequency, and then decrease sample size once there is evidence that the contractor's process is under control	Low, Low, Med, High, Med	Med, Low, Med, Med, High	High, High, Med, Low, High	Yes, No, Maybe, Yes, Maybe
3	Randomize the AQCs to be tested at any one location (i.e., do not test all AQCs at all locations)	Med, Med, Med, Med, Med	Med, Med, High, High, Med	High, Med, Med, Low, Med	Yes, Yes, Maybe, Maybe, Yes
4	Reduce sample size to 3 per lot	Low, Med, Med, High, Med	Med, Med, High, Med, High	High, High, High, High, High	Maybe, Maybe, Maybe, No, No
5	Reduce or eliminate the averaging of multiple samples	Low, Low, Med, Med, Low	Low, Med, Med, Med, Med	Med, High, High, High, Med	Maybe, No, Yes, No, Maybe
6	Use third–party testing for acceptance (e.g. by commercial lab representing the agency)	Low-Med, High, High, High, High	Low, Med, Low, High, Med	High, High, High, High, Med	Yes, Maybe, Yes, No, Maybe
7	Use automated equipment and plant records to replace/decrease testing of asphalt content, gradation, air content, strength, etc.	Low, Med, Med, High, Med	Low, High, Low, Med, high	Med, Med, High, Med, High	Yes, Maybe, Yes, Yes, Yes
8	Test contractors with "A" ratings at a lower frequency than contractors with "C" ratings. Contractor ratings are for quality management purposes only, with no effect on bidding.	Med, Med, High, Med, Med	Low-Med, Low, Low, Med, Low	Med, Med, Low, Low, Med	Yes, Yes, Maybe, Maybe, Yes
9	Slow the project down to give agency more time to run tests	Low, Med, Low, Med, Low	Low, Low, Low, Low, Med	Med, High, Low, Low, Low	Maybe, No, No, No, No
10	Use larger lot sizes	Low-Med, Med, High, High, High	Med, Med, High, High, Low	High, High, High, High, High	Maybe, Yes, Maybe, Maybe, Yes
11	Use warranties	Med, Med, High, High, High	Low-Med, Med, Low, Low, Low	High, High, Med, High, High	Yes, Yes, Yes, Yes, Yes

Figure B-1: Panel Member Responses to Workload Reducing Alternatives

- Appropriate for State DOTs that are using CAT, but are looking for ways to improve it.

No.	Alternatives to CAT	Potential for Reducing Agency's workload	Potential for Increasing Agency's Risk of Accepting Poor Quality Products	Ease of Implementation	Worthy of Further Investigation?
		(Low, Med., High)	(Low, Med.,High)	(Low, Med.,High)	(Yes, Maybe, No)
1	Require contractor's testing staff to report to a separate unit within the contractor's organization (i.e., require a separation between the contractor's quality management team and project management team)	Low, Low, Med	Low, Low, Low	Med, Med, Med	Yes, Yes, Yes
2	Require certain certification and/or training of the contractor's technicians who perform CAT	Low, Low, Med	Low, Low, Low	Med, Med, High	Yes, Yes, yes
3	Eliminate or reduce bonuses to decrease the potential for fraud	Low, Low, Low	Low, Low, Med	High, Med, High	Yes, Yes, Maybe
4	Use larger lots to compare contractor vs. agency test results; F and t tests would have larger *n* and thus be more discerning (conditioned on normality of data)	Low, Low, Low	Low, Low, Low	High, Med, High	Yes, Yes, yes
5	Develop guidelines for applying F and t tests for CAT (e.g., what's an acceptable level of normality?)	Low, Low, Low	Low, Low, Low	Med, Med, Med	Yes, Yes, Maybe
6	Combine contractor and agency test results	Low, Med, High	Med, Med, High	Med, Med, High	Maybe, Yes, No
7	Make no changes to current CAT practices	Low, Low, Low	Low, Low, Low	High, High, High	Maybe, No, Maybe

Figure B-2: Panel Member Responses to Alternatives Suggested as Improvement to CAT

The points scored by each alternative under each category are shown in following Figures.

- Appropriate for states that are looking to reduce the agency's workload, but are not interested in using CAT.

No.	Alternatives to CAT	Potential for Reducing Agency's workload	Potential for Increasing Agency's Risk of Accepting Poor Quality Products	Ease of Implementation	Worthy of Further Investigation?	Average Rating
		(Low, Med., High)	(Low, Med.,High)	(Low, Med.,High)	(Yes, Maybe, No)	
1	Start project with normal test frequency, and then increase the lot size once there is evidence that the contractor's process is under control (Florida's approach)	1.9	2.6	2.2	2.6	2.2
2	Start project with normal test frequency, and then decrease sample size once there is evidence that the contractor's process is under control	1.8	2	2.4	2.2	2.1
3	Randomize the AQCs to be tested at any one location (i.e., do not test all AQCs at all locations)	2	1.6	2	2.6	1.9
4	Reduce sample size to 3 per lot	2	1.6	3	1.6	2.2
5	Reduce or eliminate the averaging of multiple samples	1.4	2.2	2.6	1.8	2.1
6	Use third–party testing for acceptance (e.g. by commercial lab representing the agency)	2.7	2.2	2.8	2.4	2.6
7	Use automated equipment and plant records to replace/decrease testing of asphalt content, gradation, air content, strength, etc.	2	2	2.8	2.8	2.3
8	Test contractors with "A" ratings at a lower frequency than contractors with "C" ratings. Contractor ratings are for quality management purposes only, with no effect on bidding.	2.2	2.7	1.6	2.6	2.2
9	Slow the project down to give agency more time to run tests	1.4	2.8	1.6	1.2	1.9
10	Use larger lot sizes	2.5	1.8	3	2.4	2.4
11	Use warranties	2.6	2.7	2.8	3	2.7

Figure B-3: Score of Each Alternative Suggested for Workload Reduction

- Appropriate for State DOTs that are using CAT, but are looking for ways to improve it.

No.	Alternatives to CAT	Potential for Reducing Agency's workload (Low, Med., High)	Potential for Increasing Agency's Risk of Accepting Poor Quality Products (Low, Med.,High)	Ease of Implementation (Low, Med.,High)	Worthy of Further Investigation? (Yes, Maybe, No)	Average Rating
1	Require contractor's testing staff to report to a separate unit within the contractor's organization (i.e., require a separation between the contractor's quality management team and project management team)	1.3	3.0	2.0	3.0	2.1
2	Require certain certification and/or training of the contractor's technicians who perform CAT	1.3	3.0	2.3	3.0	2.2
3	Eliminate or reduce bonuses to decrease the potential for fraud	1.0	2.7	2.7	2.7	2.1
4	Use larger lots to compare contractor vs. agency test results; F and t tests would have larger *n* and thus be more discerning (conditioned on normality of data)	1.0	3.0	2.7	3.0	2.2
5	Develop guidelines for applying F and t tests for CAT (e.g., what's an acceptable level of normality?)	1.0	3.0	2.0	2.7	2.0
6	Combine contractor and agency test results	2.0	1.7	2.3	2.0	2.0
7	Make no changes to current CAT practices	1.0	3.0	3.0	1.7	2.3

Figure B-4: Score of Each Alternative Suggested as Improvement to CAT

APPENDIX C: CONSTANT for SKIP-LOT SAMPLING PLAN

Tables C-1 through C-11 obtained from Perry 1970 (*21*).

Table C-1: Constants for SkSP-2 Plans Having Acceptance Constant of 0

OR	***f***	***I***	**$np_{.95}$**
44.891	1	-	0.051
32.000	2/3	4	0.072
33.377	″	6	0.069
34.373	″	8	0.067
34.894	″	10	0.066
35.984	″	12	0.064
36.566	″	14	0.063
25.888	1/2	4	0.089
27.417	″	6	0.084
28.788	″	8	0.080
29.909	″	10	0.077
31.122	″	12	0.074
32.437	″	14	0.071
19.370	1/3	4	0.119
21.324	″	6	0.108
23.030	″	8	0.100
24.500	″	10	0.094
25.876	″	12	0.089
27.094	″	14	0.085
15.903	1/4	4	0.145
17.992	″	6	0.128
19.853	″	8	0.116
21.523	″	10	0.107
22.802	″	12	0.101

24.242	″	14	0.095
13.814	1/5	4	0.167
15.883	″	6	0.145
17.715	″	8	0.130
19.353	″	10	0.119
20.936	″	12	0.110
22.360	″	14	0.103

- Single Sampling Reference Plan

Table C-2: Constants for SkSP-2 Plans Having Acceptance Constant of 1

OR	***f***	***I***	**np$_{.95}$**
10.946	1	-	0.355
9.112	2/3	4	0.427
9.284	″	6	0.419
9.442	″	8	0.412
9.581	″	10	0.406
9.701	″	12	0.401
9.823	″	14	0.396
8.056	1/2	4	0.483
8.330	″	6	0.467
8.568	″	8	0.454
8.781	″	10	0.443
8.963	″	12	0.434
9.131	″	14	0.426
6.816	1/3	4	0.571
7.204	″	6	0.540
7.539	″	8	0.516
7.827	″	10	0.497
8.071	″	12	0.482

8.294	″	14	0.469
6.103	1/4	4	0.638
6.549	″	6	0.594
6.922	″	8	0.562
7.244	″	10	0.537
7.524	″	12	0.517
7.780	″	14	0.500
5.629	1/5	4	0.692
6.107	″	6	0.637
6.505	″	8	0.598
6.849	″	10	0.568
7.151	″	12	0.544
7.424	″	14	0.524

- Single Sampling Reference Plan

Table C-3: Constants for SkSP-2 Plans Having Acceptance Constant of 2

OR	***f***	***I***	**$np_{.95}$**
6.509	1	-	0.818
5.687	2/3	4	0.936
5.767	″	6	0.923
5.843	″	8	0.911
5.901	″	10	0.902
5.960	″	12	0.893
6.008	″	14	0.886
5.189	1/2	4	1.026
5.323	″	6	1.000
5.437	″	8	0.979
5.533	″	10	0.962

5.621	″	12	0.947
5.699	″	14	0.934
4.587	1/3	4	1.161
4.778	″	6	1.114
4.938	″	8	1.078
5.079	″	10	1.048
5.198	″	12	1.024
5.307	″	14	1.003
4.218	1/4	4	1.263
4.447	″	6	1.197
4.637	″	8	1.148
4.795	″	10	1.110
4.938	″	12	1.078
5.060	″	14	1.052
3.964	1/5	4	1.344
4.218	″	6	1.262
4.421	″	8	1.204
4.597	″	10	1.158
4.748	″	12	1.121
4.883	″	14	1.090

- Single Sampling Reference Plan

Table C-4: Constants for SkSP-2 Plans Having Acceptance Constant of 3

OR	***f***	***I***	**$np_{.95}$**
4.890	1	-	1.366
4.382	2/3	4	1.525
4.430	″	6	1.508
4.475	″	8	1.493

4.514	″	10	1.480
4.551	″	12	1.468
4.582	″	14	1.458
4.063	1/2	4	1.645
4.147	″	6	1.611
4.220	″	8	1.583
4.283	″	10	1.560
4.338	″	12	1.540
4.387	″	14	1.523
3.668	1/3	4	1.822
3.796	″	6	1.760
3.900	″	8	1.713
3.991	″	10	1.674
4.069	″	12	1.642
4.137	″	14	1.615
3.425	1/4	4	1.952
3.577	″	6	1.868
3.701	″	8	1.805
3.807	″	10	1.755
3.898	″	12	1.714
3.979	″	14	1.679
3.252	1/5	4	2.056
3.423	″	6	1.952
3.561	″	8	1.876
3.677	″	10	1.817
3.777	″	12	1.769
3.864	″	14	1.729

- Single Sampling Reference Plan

Table C-5: Constants for SkSP-2 Plans Having Acceptance Constant of 4

OR	***f***	***i***	**np.95**
4.057	1	-	1.970
3.691	2/3	4	2.166
3.729	″	6	2.144
3.760	″	8	2.126
3.789	″	10	2.110
3.814	″	12	2.096
3.838	″	14	2.083
3.460	1/2	4	2.311
3.522	″	6	2.270
3.574	″	8	2.237
3.620	″	10	2.208
3.660	″	12	2.184
3.696	″	14	2.163
3.167	1/3	4	2.525
3.262	″	6	2.451
3.339	″	8	2.394
3.406	″	10	2.347
3.464	″	12	2.308
3.514	″	14	2.275
2.984	1/4	4	2.681
3.097	″	6	2.581
3.191	″	8	2.505
3.271	″	10	2.444
3.338	″	12	2.395
3.397	″	14	2.353
2.285	1/5	4	2.805

2.982	″	6	2.681
3.085	″	8	2.591
3.172	″	10	2.520
3.247	″	12	2.462
3.312	″	14	2.414

- Single Sampling Reference Plan

Table C-6: Constants for SkSP-2 Plans Having Acceptance Constant of 5

OR	***f***	***i***	**np.95**
3.549	1	-	2.613
3.264	2/3	4	2.842
3.293	″	6	2.817
3.318	″	8	2.795
3.341	″	10	2.776
3.361	″	12	2.760
3.378	″	14	2.746
3.080	1/2	4	3.012
3.129	″	6	2.964
3.171	″	8	2.925
3.207	″	10	2.892
3.240	″	12	2.863
3.267	″	14	2.839
2.847	1/3	4	3.259
2.923	″	6	3.173
2.985	″	8	3.107
3.038	″	10	3.053
3.084	″	12	3.007
3.124	″	14	2.969
2.699	1/4	4	3.438

2.792	″	6	3.322
2.867	″	8	3.235
2.930	″	10	3.166
2.984	″	12	3.108
3.031	″	14	3.060
2.593	1/5	4	3.579
2.698	″	6	3.438
2.782	″	8	3.334
2.852	″	10	3.252
2.911	″	12	3.186
2.963	″	14	3.130

- Single Sampling Reference Plan

Table C-7: Constants for SkSP-2 Plans Having Acceptance Constant of 6

OR	***f***	***i***	**np.95**
3.206	1	-	3.285
2.971	2/3	4	3.545
2.995	″	6	3.517
3.016	″	8	3.492
3.346	″	10	3.471
3.051	″	12	3.452
3.065	″	14	3.436
2.819	1/2	4	3.737
2.860	″	6	3.683
2.894	″	8	3.639
2.925	″	10	3.601
2.951	″	12	3.569
2.974	″	14	3.542
2.624	1/3	4	4.015

2.688	″	6	3.919
2.740	″	8	3.844
2.784	″	10	3.783
2.822	″	12	3.732
2.855	″	14	3.689
2.500	1/4	4	4.216
2.578	″	6	4.086
2.641	″	8	3.988
2.694	″	10	3.910
2.739	″	12	3.846
2.778	″	14	3.791
2.410	1/5	4	4.373
2.499	″	6	4.215
2.570	″	8	4.099
2.629	″	10	4.007
2.678	″	12	3.933
2.722	″	14	3.870

- Single Sampling Reference Plan

Table C-8: Constants for SkSP-2 Plans Having Acceptance Constant of 7

OR	***f***	***i***	**$np_{.95}$**
2.957	1	-	3.981
2.757	2/3	4	4.270
2.777	″	6	4.238
2.795	″	8	4.211
2.811	″	10	4.187
2.825	″	12	4.167
2.837	″	14	4.149

2.627	1/2	4	4.482
2.662	″	6	4.422
2.692	″	8	4.373
2.717	″	10	4.332
2.740	″	12	4.296
2.759	″	14	4.266
2.459	1/3	4	4.788
2.514	″	6	4.682
2.559	″	8	4.600
2.597	″	10	4.533
2.630	″	12	4.476
2.658	″	14	4.428
2.351	1/4	4	5.009
2.419	″	6	4.867
2.473	″	8	4.759
2.519	″	10	4.673
2.558	″	12	4.602
2.592	″	14	4.542
2.273	1/5	4	5.182
2.350	″	6	5.009
2.412	″	8	4.881
2.463	″	10	4.780
2.506	″	12	4.698
2.543	″	14	4.629

- Single Sampling Reference Plan

Table C-9: Constants for SkSP-2 Plans Having Acceptance Constant of 8

OR	***f***	***i***	**$np_{.95}$**
2.768	1	-	4.695
2.593	2/3	4	5.011

2.611	″	6	4.977
2.627	″	8	4.947
2.641	″	10	4.921
2.653	″	12	4.899
2.663	″	14	4.879
2.480	1/2	4	5.242
2.510	″	6	5.178
2.536	″	8	5.124
2.559	″	10	5.079
2.578	″	12	5.040
2.595	″	14	5.007
2.331	1/3	4	5.576
2.380	″	6	5.461
2.419	″	8	5.371
2.453	″	10	5.298
2.481	″	12	5.237
2.507	″	14	5.184
2.235	1/4	4	5.816
2.296	″	6	5.661
2.344	″	8	5.544
2.384	″	10	5.451
2.419	″	12	5.373
2.448	″	14	5.308
2.166	1/5	4	6.002
2.235	″	6	5.815
2.289	″	8	5.676
2.334	″	10	5.567
2.372	″	12	5.478
2.405	″	14	5.403

- Single Sampling Reference Plan

Table C-10: Constants for SkSP-2 Plans Having Acceptance Constant of 9

OR	***f***	***i***	**$np_{.95}$**
2.618	1	-	5.425
2.464	2/3	4	5.676
2.479	″	6	5.730
2.493	″	8	5.698
2.505	″	10	5.670
2.516	″	12	5.646
2.526	″	14	5.624
2.361	1/2	4	6.017
2.389	″	6	5.947
2.412	″	8	5.889
2.433	″	10	5.840
2.450	″	12	5.799
2.465	″	14	5.762
2.229	1/3	4	6.375
2.272	″	6	6.252
2.308	″	8	6.155
2.338	″	10	6.077
2.364	″	12	6.010
2.386	″	14	5.954
2.143	1/4	4	6.633
2.197	″	6	6.467
2.240	″	8	6.341
2.276	″	10	6.241
2.307	″	12	6.158
2.334	″	14	6.087
2.080	1/5	4	6.883

2.142	″	6	6.632
2.191	″	8	6.484
2.232	″	10	6.366
2.266	″	12	6.270
2.295	″	14	6.189

- Single Sampling Reference Plan

Table C-11: Constants for SkSP-2 Plans Having Acceptance Constant of 10

OR	***f***	***i***	**$np_{.95}$**
2.497	1	-	6.169
2.358	2/3	4	6.535
2.372	″	6	6.495
2.385	″	8	6.461
2.396	″	10	6.431
2.405	″	12	6.405
2.414	″	14	6.382
2.265	1/2	4	6.802
2.290	″	6	6.727
2.311	″	8	6.666
2.329	″	10	6.614
2.345	″	12	6.569
2.359	″	14	6.530
2.145	1/3	4	7.185
2.184	″	6	7.053
2.217	″	8	6.950
2.244	″	10	6.866
2.267	″	12	6.796
2.287	″	14	6.735
2.066	1/4	4	7.459

2.115	″	6	7.283
2.155	″	8	7.149
2.188	″	10	7.042
2.216	″	12	6.953
2.240	″	14	6.878
2.009	1/5	4	7.673
2.066	″	6	7.459
2.110	″	8	7.300
2.147	″	10	7.175
2.178	″	12	7.073
2.025	″	14	6.987

- Single Sampling Reference Plan

VITA

Name:	Sujay Sudhir Wani
Address:	3108 Gables Drive NE, Atlanta, GA-30319
Email Address:	sujaywani@gmail.com

Education:	B.E., Civil Engineering, Veermata Jijabai Technological Institute, Mumbai, India, 2007
	M.S., Civil Engineering, Texas A&M University, 3136 TAMU College Station, Texas, USA