QUALITY ASSURANCE MONITORING & SAMPLING METHOD DEVELOPMENT FOR ITS AND TRAFFIC CONTROL SIGNAL DEVICES

Contract BD543-18

Dr. Jim Simpson
Florida State University Department of Industrial and Manufacturing Engineering

For
The Florida Department of Transportation
Traffic Engineering and Operations Office

Final Report - March 2008

Final Report

Reporting Period: July 2007-March 2008

Reference: *Executed contract- Federal Aid Project CFDA-20.205 / Contract BD543-18, "QUALITY ASSURANCE MONITORING & SAMPLING METHOD DEVELOPMENT FOR ITS AND TRAFFIC CONTROL SIGNAL DEVICES."*

Investigators:

Faculty

Jim Simpson (PI) – Phone: (850) 410-6354 / Email: simpson@eng.fsu.edu

Department of Industrial and Manufacturing Engineering
FAMU/FSU College of Engineering, Florida State University
2525 Pottsdamer Street, Tallahassee, Fl 32310

Student Assistants

Current
Graduate – Aaron Martin

Past
Graduate – Michael Lisansky

FDOT Project Manager:

Jeffrey M. Morgan
jeffrey.morgan@dot.state.fl.us

Traffic Engineering Research Laboratory
Traffic Engineering and Operations Office
Florida Department of Transportation
605 Suwannee Street, MS 36
Tallahassee, FL 32399-0450

Technical Report Documentation Page

<table>
<tr><td>1. Report No.</td><td>2. Government Accession No.</td><td>3. Recipient’s Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
QUALITY ASSURANCE MONITORING & SAMPLING METHOD DEVELOPMENT FOR ITS AND TRAFFIC CONTROL SIGNAL DEVICES</td><td>5. Report Date:
2008/3/31 (updated 05/15/2008)</td></tr>
<tr><td>6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Authors:
Dr. Jim Simpson (PI)
Aaron Martin (Research Assistant)</td><td>8. Performing Organization Report No.</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization:
Department of Industrial and Manufacturing Engineering
FAMU-FSU College of Engineering, Florida State University
2525 Pottsdamer Street, Tallahassee, Fl 32310</td><td>10. Work Unit No. (TRAIS)</td></tr>
<tr><td>11. Contract or Grant No.
BD543-18</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address
The Florida Department of Transportation
Research Center
605 Suwannee Street, MS 30
Tallahassee, Fl 32399-0450</td><td>13. Final Report
Start Date: 2007/5/1
End Date: 2008/3/31</td></tr>
<tr><td>14. Sponsoring Agency Code</td></tr>
<tr><td colspan="3">15. Supplementary Notes:
Additional reviewers include:
Jeffrey Morgan (FDOT Project Manager), Steven Bentz (FDOT), Sivam Ramalingam (FDOT)</td></tr>
<tr><td colspan="3">16. Abstract:
Florida’s traveling public expects safe, uniform and efficient signalized intersections and intelligent transportation systems as they travel the highways of Florida. Millions of dollars are required to effectively deploy and maintain these systems and it is a goal of the Florida Department of Transportation (FDOT) to ensure end users of these transportation systems receive their full benefit. To help achieve this goal the FDOT developed an approved product list (APL) that lists transportation devices that have been tested and verified to meet FDOT standards, are reliable, and meet the highest possible quality standards. To ensure the APL contains only product that helps to achieve this goal, a systematic monitoring and sampling procedure must be in place that filters out those devices that do not meet state requirements. Such a procedure should include testing of APL devices before and after installation, modern quality assurance and quality control methods, and improved communication between FDOT personnel that maintain the APL and end users of transportation products listed on the APL.</td></tr>
<tr><td colspan="3">17. Key Words:
Quality Assurance, Highway Safety, Traffic Control, Traffic Control Device, Traffic Signals, Standardization, Transportation Planning, Field Tests, Transportation Systems, Intelligent Transportation Systems, Monitoring, Sampling, Research Projects</td></tr>
<tr><td colspan="3">18. Distribution Statement</td></tr>
<tr><td>19. Security Classification</td><td>20. Security Classification</td><td>21. Number of Pages: 51</td></tr>
<tr><td colspan="3">22. Price</td></tr>
</table>

Disclaimer Page

(1) "The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the state of Florida Department of Transportation or the U.S. Department of Transportation."
(2) "Prepared in cooperation with the state of Florida Department of Transportation and the U.S. Department of Transportation."

QUALITY ASSURANCE MONITORING & SAMPLING METHOD DEVELOPMENT FOR ITS AND TRAFFIC CONTROL SIGNAL DEVICES

BACKGROUND

Florida's traveling public expects safe, uniform and efficient signalized intersections and intelligent transportation systems as they travel the highways of Florida. Millions of dollars are required to effectively deploy and maintain these systems and it is a goal of the Florida Department of Transportation (FDOT) to ensure end users of these transportation systems receive their full benefit. To help achieve this goal the FDOT developed an approved product list (APL) that lists transportation devices that have been tested and verified to meet FDOT standards, are reliable, and meet the highest possible quality standards. To ensure the APL contains only product that helps to achieve this goal, a systematic monitoring and sampling procedure must be in place that filters out those devices that do not meet state requirements. Such a procedure should include testing of APL devices before and after installation, modern quality assurance and quality control methods, and improved communication between FDOT personnel that maintain the APL and end users of transportation products listed on the APL.

OBJECTIVES

The primary purpose of this research is to continuously improve the utility of the FDOT's product approval program by gathering information, improving communication, educating customers, and promoting operations at the FDOT Traffic Engineering

Research Laboratory (TERL). To gather information and improve communication, a series of interviews conducted in various districts of the state are required. These interviews are opportunities to glean information regarding the current level of quality assurance being emphasized in the field and improve such practices. They will also serve as opportunities to improve the communication link between TERL personnel and end users by means of online forms and phone calls.

This improved communication will likely lead to concerns regarding nonconforming vendors. To further investigate such distress calls (or even to actively seek problem devices), a statistically sound sampling method is needed to provide the level of confidence needed to ensure devices placed in the field are of consistently high quality, operate well within their life-cycle, and are compliant to FDOT specifications. If implemented, the quality assurance field monitoring method will be deployed to maintain durable, safe, and reliable products in the state of Florida.

FINDINGS AND CONCLUSIONS

The interviews with end users conducted throughout this project have yielded similar results. In general, traffic devices listed on the APL and purchased by end users are of high quality. Few cases have been recorded in which this is not the case. While statistical methods are not being actively deployed to measure the reliability of products currently being used in the field, it is not uncommon to witness such techniques being used to measure device's quality before installation. This is particularly true in District Three. It has also been found that, in general, end users are comfortable in using the APL web site to report problems with quality, although awareness of the site could be increased throughout the state. However, these findings come from two of the best maintaining agencies in the panhandle area. To achieve a more well rounded and accurate portrayal of how things really are, it is strongly recommended that further interviews be conducted with maintaining agencies from areas with different population densities and levels of maturation.

Furthermore, these interviews strengthened the communication link between the TERL and the maintaining agencies of Florida. One of the benefits of such a communication link is the ability of TERL personnel to receive feedback from end users of APL devices. Valuable recommendations have been gathered from this feedback. For example, end users feel as though a web site dedicated solely to maintaining agencies would be a great way to foster communication between end users and TERL personnel. Such a web site would be a prime medium for posting technical bulletins regarding the level of quality of APL devices in different parts of the state. It would also be a great opportunity to develop an online forum in which maintaining agencies can converse about the APL and quality related topics. Links to device complaint forms and the APL could also be posted on the web site.

APL Manufacturer Facility Audit

In addition to interviewing end users, the author audited a vendor throughout the course of this research to further investigate the level of quality of APL devices in the state. The audit involved a document review session and a manufacturing facility tour. The documents strongly indicated that the company was dedicated to an effective QA system. The manufacturing facility tour reinforced this and demonstrated the QA system in action.

Statistical Sampling

Various statistically sound sampling procedures for sampling individual APL devices have been studied and proposed to further investigate claims of nonconforming product from end users or to actively seek such problems. These candidate techniques include hypothesis testing, acceptance sampling (such as single, double, and sequential acceptance sampling), Wald's sequential probability ratio test (SPRT), confidence interval construction, and statistical methods for reliability data. However, in most cases, acquiring the desired level of confidence in such statistical methods involved the collection of too large a number of devices to be economical. For devices such as hardware, signal brackets, etc., larger sample sizes may be feasible, but for more expensive devices such as traffic controllers or dynamic message signs, large sample sizes are impossible. Because of this, these statistical sampling strategies will not be feasible for field sampling of individual devices.

Intersection Sampling

An alternative to device sampling is intersection sampling. This idea involves examining intersections as whole and recording defects at a system level rather than an individual device level. Logs at intersections should be kept that record installation dates, problems, failures, and repairs of devices. This practice is already being done by some end users (such as Bay County); the next step would be to relay the information back to TERL personnel in an organized way. Another alternative to field sampling is to sample devices prior to installation. If deployed successfully, such alternatives to device sampling would potentially enable TERL personnel to make use of the aforementioned statistical sampling methods. For this to work, end users would need to follow a set of best practices. For example, end users should document tests that are performed prior to installation. Specific details should be recorded including what devices were tested, how many in total were tested, how many passed the test, how many failed the test, and what types of failures were observed. The results should then be relayed to TERL staff and used to confidently use the statistical methods outlined in the body of this report.

BENEFITS

The above research activities support the FDOT's mandate to ensure a uniform and compliant transportation system as required by Section 316.0745, Florida Statutes.

The findings of this report will aid TERL personnel in furthering an effective QA system. Communication with end users and TERL personnel has increased (particularly in District 3) since the onset of this project. Notifications from end users will signal the maintainers of the APL to investigate nonconforming products. The statistical methods outlined are ideal for assessing the validity of claims from end users to discipline vendors and filter out nonconforming products or to assess the level of quality of APL devices of interest. By doing so, the traveling public of Florida will enjoy improved transportation safety, uniformity and efficiency.

This research project was conducted by Dr. Jim Simpson, of the FSU/ FAMU College of Engineering. For more information, contact Jeffrey M. Morgan, FDOT Project Manager, at:

Phone: 850-921-7354

Email: Jeffrey.Morgan@dot.state.fl.us

Table of Contents

Chapter One: Introduction 1
Chapter Two: Current Quality Analysis 3
  Interview with the End Users of Traffic Devices 3
    Interview Goals 3
    Interview Introduction 4
    Questions 4
    Post Interview Discussion 6
  Tallahassee Interviews 6
    General Quality Level 6
    APL Issues 7
    Communication with Central Office 7
    Miscellaneous Comments 8
  District Three (Bay County) 10
    General Quality Level 10
    APL Issues 11
    Communication with Central Office 11
    Miscellaneous Comments 11
  APL Vendor Audit 12
    Document Review .... **Error! Bookmark not defined.**
    Management Review Report of QA System 12
    Quality Manual 13
    Latest Audit Report by Third Party and Internal Audit Report **Error! Bookmark not defined.**
    Non-Conformance Reports for Products and Processes .... **Error! Bookmark not defined.**
    Test Equipment and Calibration Reports .... **Error! Bookmark not defined.**
    Supplier / Vendor Audits 13
    Plant Tour 13
Chapter Three: Statistically Sound Sampling 14
  Considerations for Statistical Sampling 14
    Device Grouping 14
  Hypothesis Testing 16
    Hypothetical Example of a Hypothesis Test on a Proportion 17
  Acceptance Sampling 20
  Introduction to Sequential Experiments 22
    Properties of the Curtailed Inspection Plan 24
  Conducting the Sequential Probability Ratio Test 26
    Properties of the Sequential Probability Ratio Test 27
    Automating the SPRT with R 28
  Confidence Interval on a Proportion 30
  Intersection Sampling 32
  Reliability Engineering Methods 33
    Censoring 34

Introductory Reliability Statistics ... 35
Nonparametric Estimation ... 36
Parametric Models for Reliability Data ... 37
Chapter Four: Conclusion ... 41
References ... 45
Appendix 1: R Program ... 46
Appendix 2: Acronym List ... 48

Chapter One: Introduction

The Florida Department of Transportation (FDOT) Traffic Engineering Research Lab (TERL) is conducting research to ensure that transportation devices contained on the FDOT Approved Product List (APL) and purchased by the state of Florida meet FDOT standards and specifications and are of the highest possible quality. The two primary goals of this project are to (1) investigate current field sampling/inspection requirements to determine the current level of quality assurance being emphasized in the field and (2) develop a statistically sound sampling method that will provide the level of confidence needed to ensure devices placed in the field are of consistently high quality, operate well within their life-cycle, and are compliant to FDOT's specifications [15].

To achieve the first part of this goal, the researchers and TERL personnel conducted interviews with the end users of transportation devices listed on the APL. Results from these interviews are presented in this paper, but in short they strengthened the communication link between end users and TERL personnel. This fosters the flow of information regarding problems with devices in the field.

Research regarding the development of the statistical sampling plan has yielded numerous possibilities, although limited resources may inhibit them from being implemented. Some of these alternatives include standard hypothesis testing, acceptance sampling, sequential acceptance sampling, confidence interval construction, and statistical methods for reliability data. If implemented, the quality assurance field monitoring method will be deployed to actively seek problem devices and assess the level of reliability of devices on the APL. Consequently, more durable, safe, and dependable products will reach the end users of Florida.

In short, the primary purpose of this research is to continuously improve the utility of the FDOT's product approval program by gathering information, improving communication, educating customers, and promoting operations at the TERL.

Chapter Two: Current Quality Analysis

To examine the current level of quality in the field, interviews with maintaining agencies were completed as well as an APL manufacturer plant audit. First, this section describes the interviewing process from start to finish. The planning phase is described first. Then a summary of the interviews are presented with a description of what valuable information can be gleaned from them. Also, a summary of an APL manufacturer plant audit is provided.

Interview with the End Users of Traffic Devices

The interviews were done by TERL staff and research personnel from the current research contract. The TERL staff and research personnel predetermined the necessary questions to be asked to satisfy the goals of the interview. The interviews were expected to take place once, although a periodic, long-term arrangement may continuously assure that the communication link is strong. The goals of the interview are presented in the next subsection.

Interview Goals

- Gather information regarding questionable APL manufacturers.
- Develop a system to continue to receive information regarding problem product and vendors.
- Improve communication between headquarters, districts, and maintaining agencies.
- Discuss the development of a web page for the maintaining agencies.
- Educate end users about:
    - APL complaint reporting system
    - APL complaint resolution process

- Promote the operations at the TERL as being capable of minimizing time, money, and manpower required to maintain or replace problem products.

Interview Introduction

This stage of the conversation served as a quick explanatory discussion to inform the interviewee of our intentions for conducting the interview. First, background information was given to the interviewee regarding what the interviewer's background and research goals are. The interviewer explained the work of the TERL such as determining which vendors have high enough quality standards and meet FDOT device specifications to be placed on the Approved Product List, or APL. The APL is a list of transportation products that have been verified by the TERL to be in conformance with FDOT specifications, and manufactured by vendors that have proven they have quality systems with standards high enough to produce consistent quality products.

To further measure the level of quality of transportation devices, it was explained that strengthening the communication link between the Central Office, the districts, and the maintaining agencies was of utmost importance. It was also made clear that the interviewer was asking questions to help maintaining agency personnel reduce (1) the amount of time spent on maintenance calls, (2) the amount of money spent on replacement and repair, and (3) the number of devices that are used that do not uphold the high level of quality to remain on the APL.

Questions

1.) Do you have any problem vendors or devices?

2.) If so, who are they? What are they?

3.) If so, documented evidence will strengthen our case to drive the problem vendor to increase the level of reliability and quality. Do you have any documented evidence or failed product?

- At this point, we should take any failed product that they have as additional proof to the vendor in question.

4.) Do you make use of the Certified Traffic Control Signal Device Complaint Form?

5.) What about the form named in (4) above makes it easy and attractive to use?

6.) What about the form named in (4) above makes it burdensome and undesirable to use?

7.) Is there another method for reporting device malfunctions that you would prefer?

- At this point, we should give the interviewee the link to the FDOT website to fill out the complaint form named in (4) above.

8.) How do you use the APL?

9.) If it seems as though the interviewee is not using the APL, we should educate them on its use.

10.) Have you ever registered a complaint about poor product with FDOT? If so, what was the response you received?

11.) Do you have any comments, concerns, or complaints regarding the APL or the overall product approval program?

12.) How do you make sure that the products you are buying work well? (e.g.- do you perform tests to measure their quality and reliability?)

13.) If you use a low bid system of buying product, how do you make sure that the product you purchase meet your specifications?

14.) Do you have any records of device installation dates, manufacturer, etc?

15.) Is there an incoming shipment (receiving) area?

Post Interview Discussion

To finish the discussion, the interviewer reiterated the intent to bridge the communication gap between the districts, maintaining agencies, and the TERL in Tallahassee. In some cases, the interviewer also informed the interviewee of the plan to devise a statistical sampling method to inexpensively and effectively perform random sampling to measure the reliability of transportation devices.

City of Tallahassee Interviews

General Quality Level

Interviews with three traffic engineers employed by the City of Tallahassee (COT) were conducted to begin to assess the level of quality of APL devices used in the state of Florida. Wayne Bryan [3] handles operations at COT's Transportation Management Center, Don Dobson [4] handles COT's Signal Operations, and Randy Powell [12] handles COT's Signal Construction. The results of these three interviews were very similar. These three COT traffic engineers asserted that poor device quality is not a severe problem in Tallahassee. The only incident that was recounted involved 358 controller fiber optic modems [3]. When the problem was explained to the vendor, the company replaced the optics in all 358 fiber optic modems. Other than that specific instance, there were no specific instances of poor device quality in Tallahassee, and this incident was promptly resolved by the vendor.

COT receives these devices directly from vendors, and before they install them in intersections around Tallahassee they test each and every device. So, 100% of the devices that they install are sampled and tested, eliminating any need of random sampling of brand new devices. This was very reassuring to hear (as sampling before installation was one of our ideas for the random sampling method), and this is likely a primary reason why poor quality of devices is not a severe problem in Tallahassee.

APL Issues

In general, the consensus was that the APL web site is fairly easy to navigate. Mr. Dobson stated that he finds the web site easy to navigate, thinks the organization of the product is done well, and finding product on the list is easy. He didn't have any recommendations for making the web site more user-friendly. He also has made use of the device complaint form, stating that the form is not difficult to track down from the web site. Mr. Powell felt also that the APL web site is organized well for the increasing complexity of the list, and he also did not have any suggestions for web site improvement.

Mr. Bryan did discuss a specific problem with the web site. He stated that an improved layout of the main FDOT web site, to include easy-to-find links to the APL list and the APL device complaint form would make communication with the TERL easier. After sharing this suggestion with the project manager, Jeff Morgan (TERL), Mr. Morgan made changes to the Traffic Operations web site that incorporated Mr. Bryan's suggestion. Additional changes to the main FDOT web site will be addressed to appropriate FDOT web site personnel.

Communication with Central Office

One of the proposals of this project has been to establish a "hotline" for Florida traffic engineers to call and register complaints with APL product. None of the traffic engineers interviewed seemed to feel that a problem device hotline was necessary. They felt that Tallahassee (and most of northern Florida, such as District 3) is close knit. In other words, they feel comfortable calling Mr. Morgan, or a fellow TERL employee, Mr. Carl Morse, at the TERL because they know these personnel well and share a strong communication link with them. Another interesting find is that these engineers use the International Municipal Signal Association's IMSA online forum. This forum provides members with an avenue to check and see how well devices are performing around the rest of the state.

Based on this, Mr. Powell suggested that instead of a phone number to a hotline, an online forum, such as the IMSA forum, should be developed for traffic device maintainers. This forum should be opened to all types of traffic devices such as traffic signals and signs, speed limit signs, pavement markers, etc. Another suggestion was to open this forum up to APL vendors. Although vendors may "gloss up" their products on their forum posts, they are also a source of valuable information that should not be ignored (particularly because honesty leads to future business, and vendors will want future business).

Miscellaneous Comments

An issue that heavily influences Tallahassee's purchasing decisions is the topic of sole source. This issue becomes important when the ability of a device to be compatible with the devices that are already installed appears. For example, fiber modems of the same model are necessary to be able to transmit video data from an intersection back to the Transportation Management Center (TMC).

One major problem that Mr. Dobson mentioned was power outages caused by storms, squirrels, etc. When power is lost, a backup form of energy must be in place to power the intersection. If the backup power source is not reliable, then traffic signals may go dark, putting the safety of drivers into jeopardy.

Mr. Dobson shared a personal opinion regarding this issue. Uninterruptible Power Supply systems (UPS) is a type of battery powered backup supply that is used as a secondary power source in the case of a power outage. He feels that these battery backups do not perform well once they have been in the field for an extended period of time. He recommended installing generators at intersections instead of the UPS units because the generators perform better and longer in the event of a power outage.

Mr. Dobson also said that in some instances, problems with the software are troublesome. For example, he recently was instructed to resolve a problem with an intersection located on a one way road. He spent time and resources troubleshooting the intersection, but discovered that everything was working correctly. The problem was actually with the controller software; the program was written for two way roads and didn’t accommodate the one way road situation. He did assure me that when software is a problem, the vendors are fast acting to replace the software with corrections.

Mr. Powell shared another problem regarding contractors that occasionally substitute cheaper materials to reduce costs. Another problem appears when, after a product has been approved, the vendor may update their product with a newer version that has is not tested. To resolve this problem, he stated that it is crucial to get feedback from the field. His observation goes hand in hand with the direction of this project, as building a strong communication link between end users and TERL staff will lead to better products on the APL.

District Three (Bay County)

General Quality Level

An interview with traffic engineers from Bay County was performed. Mr. Gene Syfrett [17], a traffic engineer with FDOT District Three, was interviewed along with several other traffic engineers that work with Bay County's maintaining agency. This interview provided similar information to the interviews performed in Tallahassee. These traffic engineers stated that poor device quality is not a severe problem in Bay County. The only incident that was recounted involved signal brackets that had been snapping, and the problem was resolved by the vendor.

They receive these devices directly from vendors, and before they install them in intersections around Bay County they test each and every device [17]. So, 100% of the electronic devices that they install are sampled and tested, eliminating any need of random sampling of brand new devices. However, items such as brackets are not sampled prior to installation. This is likely a primary reason why poor quality of devices is not a severe problem in Bay County. Sampling devices before installation gives a direct measure of the quality level of brand new devices, but it does not shed light on the reliability of the devices once they have been subjected to field conditions.

The maintaining agency personnel also visit intersections throughout the district and keep detailed records of problems. In each cabinet they keep a log that contains dates, problem descriptions, problem solutions, and start / finish times for work performed on the intersection. Every 90 days they perform preventative maintenance inspections on all intersections and resolve inefficiencies before they escalate into bigger problems.

APL Issues

Mr. Syfrett felt that the APL web site was fairly easy to navigate. Most of the maintaining agency personnel stated that they don't make regular use of the APL, but they still abide by the law to purchase devices from APL certified vendors. Mr. Gerald Kearney (Bay County) said that in his experience with the APL, he found it easy to use. He didn't have any recommendations for making the APL more user-friendly. In general, these traffic engineers expressed a lot of confidence in the APL. They use the APL as a means to ensure that they are buying products that meet a high standard.

Communication with Central Office

One of the proposals of this project has been to establish a "hotline" for Florida traffic engineers to call and register complaints. None of these traffic engineers seemed to feel that a problem device hotline was necessary. Instead, they felt that a web site dedicated to the maintaining agencies would be most beneficial. Such a web site would allow them to get online and find news about product quality and even communicate with the central office. The news could be transmitted through technical bulletins. This opinion is similar to the opinions of Tallahassee traffic engineers; an online forum may be another good way to strengthen the communication link.

Miscellaneous Comments

One problem that Bay County faces is the natural environment. Located near the Gulf of Mexico, many intersections are faced with salty air from the sea that causes corrosion and decay of traffic devices. This is not a new problem, and the traffic engineers interviewed have some viable solutions. For example, they may purchase stainless steel products that hold up to the corrosive environment better than other products. After talking with Bay County, it was evident that this group was very proactive; they seek out problems and target solutions.

In the future, it is important to factor into the approach for interviewing that the maintaining agency will want to have a positive working relationship with product vendors. The attitude of the interview should be such as to not "point fingers" at the maintaining agency but rather to foster a strong bond between the TERL and the maintaining agency. In order to accomplish this, a prepared message will need to be put together that explains why the program being developed will ultimately make life easier for traffic device maintainers. It should be explained that the feedback from traffic device maintainers will be used to put pressure on vendors to raise product quality, and this improved quality will be realized in the field in the form of fewer defective items. To uphold positive relations between device maintainers and vendors, it should also be explained that the feedback from maintaining agencies will remain anonymous.

APL Vendor Audit

Throughout the course of examining the current level of quality in the state, the author also audited an APL vendor. The first part of the audit involved a document review session with the company's quality manager. After the document review session, a plant tour was conducted. In summary, the APL vendor's quality assurance and quality control practices were exemplary. More detail is provided in the sections below.

Management Review Report of QA System

This report was used to determine if management is committed toward quality assurance systems. The report contained how quality assurance systems are advocated and implemented in the plant at the process and product level. The management was fully committed towards quality assurance.

Quality Manual

The manual indicated that the company is strongly committed to providing customers with the highest quality products in the transportation systems market. The manual also stressed continual improvement and the use of advanced technologies to provide customers with solutions to their needs.

The quality program is updated and continuously improved via customer feedback, failure analysis on products through quantitative data, management reviews, internal audits, preventative action, and corrective action.

Supplier / Vendor Audits

Supplier audits are performed as necessary. Suppliers are also constantly monitored and evaluated.

Plant Tour

During the plant tour various processes were shown such as assembly tasks in work cells, incoming shipping / inspection, quality inspection, burn-in testing, and outgoing shipping / inspection.

Chapter Three: Statistically Sound Sampling

This section focuses on the development of a statistically sound sampling method. First, some general considerations for sampling are presented. What follows is a technical exploration of statistical sampling methods. Some of prospects investigated in this paper include standard hypothesis testing, acceptance sampling, sequential acceptance sampling, and confidence interval construction.

Considerations for Statistical Sampling

The nature of this situation makes statistical sampling difficult; devices are in the hands of the end users and are spread across the state of Florida. As a result, the resources required to perform such sampling dramatically increase from the TERL's point of view. Time, money, and manpower become a driving issue in the feasibility of the development of the notional sampling method. In some cases, the testing of just one device may be as high as $5,000.00 [10]. However, to place confidence in a statistical test, multiple random samples must be taken. Other considerations must be taken into account, such as device grouping and post inspection device classification.

Device Grouping

Devices are subdivided four ways.

1. First, they are categorized by the manufacturer. For example, devices may be separated as being manufactured by ADDCO and Armorcast.
2. Devices are also grouped by category. Some examples of categories include vehicle detectors and pedestrian detectors.
3. Devices are further grouped by the type of device, such as a balancers and back plates. Devices can be easily found by navigating through the APL web site. One

may notice that products may be capitalized or in lower case. The lower case products are products that have been approved by planning [1].

4. The fourth technique of device grouping is based on a product's approval level. The approval level has two subsets including certified and approved products. Both of these subsets qualify products to be listed on the APL, but what separates them is that certified products meet the definition of an official traffic control device [11].

Once devices reach the end user, one important aspect that should be considered is the device's age. As the age of a device increases, the likelihood that it may fail increases. This should be accounted for somehow in diagnosing quality. Devices of similar age may be pooled together and analyzed via the proactive sampling program. The age of a device may be discerned by locating its date code, which typically specifies the date and year of manufacture. This is of particular importance when inspecting electronic devices, such as a pedestrian count down module [11].

Once a device has been sampled and tested, it is classified as either "pass" or "fail." This method lends itself to the analysis of the *proportion* of parts that pass or fail. One type of analysis using this idea is the proportional hypothesis test. By incorporating the idea of hypothesis testing into sampling lots of devices, the idea of acceptance sampling is achieved [5]. There are multiple types of tests that may be performed on a device, and there are multiple types of defects that can be detected. Defects can generally be divided into operational and non-operational defects. Some operational defects include:

1) Structural defects occur when a device fails a fatigue/stress test.

2) Electronic defects are problems with electronic components that inhibit the device from properly functioning

Some non-operational defects include:

1) Administrative defects occur when a product is missing an identification tag, making vendor identification difficult if not impossible.
2) Chemical defects occur when a device fails a chemical test (such as a chemical composition test to test for a particular type of alloy).
3) Visual defects include poor paint jobs or poor surface finishes, but do not necessarily affect the performance of a device.

Hypothesis Testing

Hypothesis testing makes up a crucial part of statistical inference, and some form of hypothesis testing may be ideal for this problem. Statisticians test alleged theories that are either believed to be true or need to be tested for the basis of argument. In such a test, the theory is divided into two competing claims. These claims are called the null hypothesis (denoted H_0) and the alternative hypothesis (denoted H_1), and they are mutually exclusive (if one is true, the other must be false). The result of such a test is to either *reject* or *fail to reject* the null hypothesis. The hypothesis test of a proportion may be an ideal statistical test for making inference about the percentage of devices that either pass or fail.

There are two types of errors that can be encountered when performing hypothesis tests. Type 1 error occurs when the null hypothesis is rejected when it is in fact true. Type 2 error occurs when the null hypothesis is not rejected when it is in fact false. In many cases, type 1 error is considered to be more serious, so the hypothesis test procedure is adjusted to minimize the chance of type 1 error occurring. In fact, this *significance level* is often chosen by the experimenter, and is denoted as P(Type 1 error) = α. Type 1 and type 2 errors are visually displayed in Table 1.

Table 1 Visual representation of type 1 and type 2 errors

		Decision	
		Reject H_0	**Don't Reject H_0**
Truth	**H_0**	Type 1 Error	Right Decision
	H_1	Right Decision	Type 2 Error

Hypothetical Example of a Hypothesis Test on a Proportion

A traffic device company manufactures Vehicle Signal Brackets. While performing a QAR in District 3, a quality engineer's task is to draw inference about the quality of these brackets in the field. He suspects that at least 80% of the 1,000,000 total brackets would pass his inspection. To test this hypothesis, he randomly samples 100 brackets and finds that 73% of them pass his test. This sample size satisfies two basic assumptions of this hypothesis. These are that $n*p > 10$ and that $n*(1-p) > 10$ (where n is the sample size and p is the observed proportion). Based on these results, should he accept or reject his hypothesis that 80% of all of the district's brackets pass his test?

STEP 1: State the hypothesis

Null Hypothesis, H_0: $P \geq 0.80$

Alternative Hypothesis, H_1: $P < 0.80$

STEP 2: Formulate an analysis plan

A significance level must be chosen. The significance level (denoted as α) is equal to the probability of type I error occurring. For this example, let's let $\alpha = 0.05$.

STEP 3: Analyze sample data

The test statistic is denoted as Z.

$Z = (P-p) / \sqrt{((p*(1-p))/N)}$, where

P = proportion that pass from sample (0.73)

p = hypothesized value of proportion that pass (0.80)

N= sample size.

So,

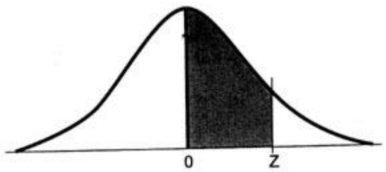
$$Z = (0.73-0.80) / \sqrt{((0.80*(0.20))/100)}$$

$$Z = -1.75$$

STEP 4: Interpret Results

In this case, the p value is equal to the probability that the z-score is less than -1.75. Using Table 2, one finds that the $Prob(z<-1.75) = 0.04$. Since the p value (0.04) is less than the significance level (0.05), we cannot accept (reject) the null hypothesis. Also, α is the prob(Type I error) = 0.05. This means that there is a 5% chance of deciding against the null hypothesis (deciding that the true proportion is not 80%) when in fact the null hypothesis is true (the true proportion really is 5%). Software was used to obtain the p-value for this example. Another method is more compatible for hand calculations. This method involves comparing the calculated Z value to a critical value. The critical value in this case is equal to $Z_{\alpha/2}$, or $Z_{0.025}$. The criteria to reject H_0 is when $|Z| > Z_{critical.}$ $Z_{critical}$= 0.065, so we reject H_0 because $|-1.75| > 0.065$.

Table 2 Areas under the standard normal curve



This table presents the area between the mean and the Z score . When Z=1.96, the shaded area is 0.4750.

Areas Under the Standard Normal Curve

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000									

Source: Adapted by permission from *Statistical Methods* by George W. Snedecor and William G. Cochran, sixth edition

Acceptance Sampling

Previous research in this area was conducted. The result of this research was a recommendation to use an acceptance sampling plan based on double sampling to aid the FDOT TERL in the reevaluation process that occurs every two years. The locations of this sampling should be random. This plan also will act as an investigation tool to check up on reported malfunctions and complaints. In this case, the locations of sampling may either be completely from the problem area or 50% from the problem area and 50% random [6].

The double sampling plan was adopted because of its ability to possibly reduce the number of samples taken. For very excellent or very inferior products, half of the necessary samples are eliminated. However, mediocre products may require the second half of the samples to further measure the product quality. Some of the disadvantages of this method are that a second trip may be necessary to gather more information, and, in some cases, the cost of the trip may greatly exceed the cost of a device test. For these situations, a single sampling plan of higher sample size should be used. This technique is flexible in that the user may choose a plan based on its ability to predict at certain probabilities of failure. If sampling and testing is difficult, then a lower sample size should be selected; likewise, as the need for accuracy increases, a larger sample size should be chosen.

Michael Lisansky [6] has also outlined a double acceptance plan to be used as a guide or as a reference. This plan is particularly useful because it lists the purported defective percentage that the user wants to test for. Then, various probabilities of accepting a lot for different sample sizes for the various nonconforming rates are listed. This information is displayed in Table 3. In this table, the columns are labeled with three numbers in parenthesis (n, c, r). Here, n is equal to the sample size, c is equal to the acceptance number (the number of detected defects such that one would accept the lot), and r is equal to the rejection number (the number of detected defects such that one would reject the lot). This table outlines possible sampling techniques for double sampling (n = 2, 5, 10, and 25) and single sampling (n = 4, 7, and 15). The probabilities in the table (the elements of the columns labeled Pa and Pa (%)) correspond to the probability of accepting a lot at the given percent defective.

Table 3 Single and Double Acceptance Sampling

Acceptance Sampling								
Fraction Defective	Double (2,0,2);(2,1,2)	Single (4,1,2)	Single (4,0,1)	Double (5,0,2);(5,1,2)	Single (7,1,2)	Double (10,0,2);(10,1,2)	Single (15,1,2)	Double (25,0,3);(25,3,4)
P	Pa	Pa	Pa	Pa	Pa	Pa	Pa	Pa
0.01	0.9996	0.999	0.961	0.9966	0.998	0.9865	0.99	0.997
0.05	0.9865	0.982	0.819	0.9309	0.951	0.7909	0.8265	0.7413
0.1	0.9525	0.938	0.67	0.7909	0.844	0.5034	0.5585	0.3729
0.15	0.9055	0.878	0.549	0.6817	0.7178	0.2995	0.3433	0.0663
0.2	0.8496	0.809	0.449	0.5034	0.592	0.1716	0.199	0.0145
0.3	0.7296	0.663	0.301	0.2995	0.3805	0.0575	0.061	0.0012
P (%)	Pa (%)	Pa (%)	Pa (%)	Pa (%)	Pa (%)	Pa (%)	Pa (%)	Pa (%)
1	99.96	99.9	96.1	99.66	99.8	98.65	99	99.7
5	98.65	98.2	81.9	93.09	95.1	79.09	82.65	74.13
10	95.25	93.8	67	79.09	84.4	50.34	55.85	37.29
15	90.55	87.8	54.9	68.17	71.78	29.95	34.33	6.63
20	84.96	80.9	44.9	50.34	59.2	17.16	19.9	1.45
30	72.96	66.3	30.1	29.95	38.05	5.75	6.1	0.12

As a hypothetical example, quality engineers are testing a product that must have less than 5 % defective. So, the row with P=5% is the row of interest in this case. The first double sampling plan ((2,0,2);(2,1,2)) consists of two possible samplings. The first sampling requires that two samples be drawn. If both of the original samples pass the test, then the sampling can stop and the lot is passed. However, there is a 98.65% chance that the lot does not conform to the 5% defective amount. If both samples fail, then the lot is rejected as nonconforming to the specified defective level of 5%. However, if one of the two original samples fails and one passes, then the second sampling must be taken.

The second sample consists of two more tests. In this case, the acceptance number is one, so if one of the two secondary samples fails, the lot is accepted. Again, the probability that this accepted lot does not conform to the 5% defective level is 98.65%.

Now, suppose that the other extreme of 25 samples is taken (the last column). Again, assume that the quality engineers decided that the product being tested must have less than 5% defective. After the first sampling, the lot is passed if none of the samples are defective. In this case, there is a 74.13% chance that the lot is accepted even though it truly may not meet the criteria of having less than 5% defective. If three of them are defective, the lot is rejected. However, if one or two are defective, then the second sampling must be taken. The second sampling consists of another 25 tests. If three of these samples fail the test, the lot is accepted as being within limits (with a 74.13% chance that the lot is actually out of limits). If four of the samples fail, the lot is rejected as nonconforming to the specified percentage of defects.

Introduction to Sequential Experiments

Random sampling is a method that is very commonly used in industry to measure the level of reliability and quality of a given product or process. However, in some cases, performing even just one sample can be very tedious and / or expensive. In such circumstances, an area of statistics known as sequential experimentation may be a useful tool to draw inference of a population while conserving the number of observations (samples drawn). The path that a sequential experiment takes depends in some way on the results that are obtained. The sequential nature of such experimentation manifests itself primarily in two ways: (1), as mentioned, the selection of the second and successive samples of the experiment depend on earlier results and (2), there must be some termination guideline that, once reached, triggers the end of the experimentation. The termination guideline is a predetermined number of defective items that is unacceptable.

Sequential experiments are useful for conducting investigations in which individual samples are drawn and determined to either pass or fail the test. This type of analysis is often referred to as acceptance sampling, and each individual sample is assumed to be independent. Typically, samples are drawn until a termination condition is met, at which point the sampling stops and the batch is either accepted as conforming or rejected as nonconforming. Consider the basic single acceptance sampling plan. In this case, a predetermined number of items (N) are tested. The rejection number, c, is the number of defective items that must be detected to reject the batch. For example, one instance of a single acceptance sampling plan may be where N=20 and c=3. 20 items are sampled, and if 3 of them are defective, the lot is rejected. If 2 or less samples are defective, the lot is accepted. Since the number of observations is fixed, any given observation is unrelated to the previous observation, and this plan is non-sequential. In other words, if the terminating condition is met after just 5 samples (n=5, c=3), the experiment would unnecessarily continue until n=N=20. The sample size is unrelated to the size of the population, because the population size is assumed uncountable or infinite.

One problem with fixing the number of observations is the possibility of expending time on observations once the termination criteria has already been met. In other words, the rejection criteria may be met and further experimentation is carried out until the predetermined number of samples (N) is drawn. One method to eliminate this type of waste is the implementation of curtailed inspection. The single sampling plan may adopt the curtailed inspection method. Inspected items are classified as effective observations (s) or defective items (r), and N becomes the total possible number of runs. Items are inspected until $r \geq c$ (in which case the batch is rejected) or until $s \geq N-c+1$ (in which case the batch is accepted). The proportion of observed defects is denoted as θ. The actual number of observations, n, can range from c to N [18]. These relevant quantities are presented in Table 4.

Table 4: Relevant quantities involved with curtailed inspection

Quantity	Description
S	Number of effective observations
R	Number of defective observations
C	Rejection number of defective observations (once c is reached, batch is rejected, experiment is terminated)
Θ	Observed proportion of defectives
N	Total possible number of observations
N	Actual number of observations

Properties of the Curtailed Inspection Plan

Some of the properties of the curtailed inspection plan are now presented. Firstly, if a batch has θ percent defective items, then the probability that the batch will be accepted given θ is given in equation (1).

$$\Pr(\text{accept I } \theta) = \sum_{r=0}^{C-1} (N! / [r!*(N-r)!])\, \theta^r (1-\theta)^{N-r}$$	(1)

This probability is deemed the operating characteristic. When plotted, it is called the Operating Characeristic (OC) curve (see Figure 1). This OC curve is the same for a sequential curtailed inspection plan and its respective non-sequential single sampling plan [16].

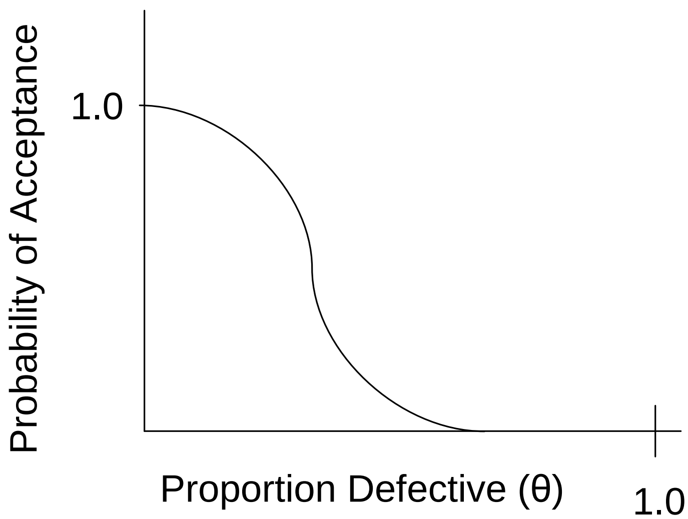


Figure 1: The OC-curve (operating characteristic curve)

The number of samples taken for the sequential curtailed sampling plan is non-constant, so the distribution of the sample size is another important property of the curtailed sampling plan. This number is estimated as the average sample number (ASN). For batches that are rejected, the sample size could be anywhere between c and N. For accepted batches, the sample size could be anywhere between (N-c+1) to N. For small values of θ, the ASN is approximately $(N - c + 1) / (1 - \theta)$. For large values of θ, the ASN is approximately c / θ.

So, the curtailed sampling plan is advantageous to the non-sequential single sampling plan in that it has the same OC curve, and for some ranges of θ it has substantially lower ASN. Wald has developed a method called the Sequential Probability Ratio Test (SPRT) that attempts to minimize the ASN for extreme values of θ (but not intermediate values) [18].

Conducting the Sequential Probability Ratio Test

The SPRT works to decide between two hypotheses; the null hypothesis is $H_0 = \theta_0$, and the alternative hypothesis is $H_1 = \theta_1$ where θ_0 and θ_1 are hypothesized percentages of defective items in a batch. If θ_0 is the true value (θ), then it should be selected with a probability of at least (1-α). On the other hand, if θ_1 is true, then it should be selected with a probability of at least (1-β). Here, α is the probability of Type I error (θ_0 is true but the test decides that θ_1 is true). Likewise, β is the probability of Type II error (θ_1 is true but the test decides that θ_0 is true). α and β are selected by the experimenter, and it is common convention to select them such that $0.01 \leq \alpha$ and $\beta \leq 0.05$. However, these probabilities can be increased to decrease the ASN (a cautionary note should be made here). Often times, the lower proportion, θ_0, is decided to as the acceptable quality level, or AQL [16]. In industry, the AQL is a level of quality that is specified and agreed to by the consumer and producer. The higher proportion, θ_1, is often set as the average outgoing quality level, or AOQL. The AOQL is the average proportion of defective items that are found via a sampling plan over a long period of time [5].

Another criterion for selecting α and β depends on the nature of the product being tested. For example, destructive crash tests in the auto industry may take more risks with higher α and β levels because the cost of each test is quite large. On the other hand, a tooth brush manufacturer would use very small α and β levels because each sample is very cheap.

Once the experimenter has chosen the appropriate proportions of defective items and significance levels (θ_0, θ_1, α, and β), the first sample can be drawn. Further sampling is continued as long as the conditions of equation (2) are satisfied. This equation contains a lower limit, the likelihood ratio, and an upper limit.

$\text{Log}\{\beta/(1-\alpha)\} < r*\text{Log}(\theta_1/\theta_0) + s*\text{Log}\{(1-\theta_1)/(1-\theta_0)\} < \text{Log}\{(1-\beta)/\alpha\}$	(2)

Sampling should be stopped if (3) is satisfied. In this case, one would fail to reject H_0 and decide for θ_0.

$r*\text{Log}(\theta_1/\theta_0) + s*\text{Log}\{(1-\theta_1)/(1-\theta_0)\} < \text{Log}\{\beta/(1-\alpha)\}$	(3)

Likewise, sampling should be stopped if (4) is satisfied. In this case, one would reject H_0 and decide for θ_1 [18]

$r*\text{Log}(\theta_1/\theta_0) + s*\text{Log}\{(1-\theta_1)/(1-\theta_0)\} > \text{Log}\{(1-\beta)/\alpha\}$	(4)

Properties of the Sequential Probability Ratio Test

The SPRT is an open sequential test. Therefore, the sample size is a random variable that takes on a skew distribution. One feature of this distribution is the ASN. The SPRT has an optimum property, but there is one major assumption that goes along with this property. Assuming that the true proportion of defective items in a population (θ) has only two possible values (θ_0 and θ_1), then the SPRT yields the lowest possible ASN at θ_0 and θ_1. Practically, this means that if an investigator selects θ_0 and θ_1 such that one of them is close to the true value of θ, the SPRT is optimum. However, at intermediate values of θ, the expected number of samples may be greater than that of a non-sequential sampling plan with the same α and β. In some cases, statisticians may truncate the SPRT at a certain sample size n_0 to conserve resources (even if the test does not terminate). However, a non-truncated SPRT will terminate with a probability of one [18].

One advantage of the SPRT is that the results for an experiment can be saved and reapplied for different values of θ_0, θ_1, α, and β. The number of defective observations (r) and the number of effective items (s) should be recorded. Then, if the investigators decide that the sample number n is growing too high, the calculations of (2) can be carried out again by reusing the same values of r and s and switching the values of θ_0, θ_1, α, and β. This may be useful in honing in on the true value of θ.

Automating the SPRT with R

R is free statistical software, and a function in Tinn-R has been written to perform the SPRT automatically [14]. It is included in the appendix. The software may be downloaded from [14] at no cost. This section presents a hypothetical example, and the relevant input / output is included.

Imagine that quality engineers at the Traffic Engineering and Research Lab are measuring the level of quality of widgets in a particular district of Florida. Samples should be drawn randomly and assumed independent. For this particular widget, it is important that absolutely no more than 30% be defective, however, 10% is much safer. In this case, the null hypothesized proportion of defective items is 0.10, and the alternative hypothesized proportion of defective items is 0.30.

The SPRT function is designed to take six arguments. They are alpha error, beta error, r, s, the null hypothesized proportion, and the alternative hypothesized proportion. Therefore, the input is of the form SPRT1(alpha, beta, r, s, prop0, prop1). The function outputs the lower limit, the likelihood ratio, the upper limit, and a recommendation to reject the null hypothesis, fail to reject the null hypothesis, or continue sampling. Hypothetically, the first sample passed the test. Assuming that $\alpha = \beta = 0.25$, the input and output in R is presented in Table 5.

Table 5 Input / output for the SPRT function in Tinn-R performing the first sample

Input	Output
SPRT1(.25, .25, 0, 1, .1, .3)	[1] "Low Limit: -0.477121" [1] "Likelihood Ratio: -0.109144" [1] "High Limit: 0.477121" [1] "More sampling required"

The likelihood ratio, -0.109144, does not exceed either of the limits; another sample must be drawn and tested. This style of sampling continues until either the null hypothesis is rejected or not. In this particular case, assume that sampling continues until r=1, s=9, and n=10. The related input / output information is presented in Table 6.

Table 6 Final input / output for the hypothetical example performed in Tinn-R

Input	Output
SPRT1(.25, .25, 1, 9, .1, .3)	[1] "Low Limit: -0.477121" [1] "Likelihood Ratio: -0.505179" [1] "High Limit: 0.477121"

	[1] "Fail to reject H0, decide true proportion is 0.100000"

Notice that the likelihood ratio is less than the lower limit. So, the conclusion is to fail to reject the null hypothesis deciding in favor of θ = 0.1. Had the likelihood ratio surpassed the upper limit in magnitude, the function would have concluded to reject the null hypothesis and choose in favor of the alternative hypothesis (θ =0.3)

Confidence Interval on a Proportion

The practice of acceptance sampling described in the sections above is intended to test whether or not lots of products are within an acceptable range of proportion defective. While this practice may apply to the necessary sampling plan under development at the TERL, it is not useful for estimating the true proportion of defective items in a population. A more suitable method for this problem is a confidence interval on a proportion. Since the result of sampling is binomial (pass, fail), then the binomial confidence interval is an ideal method for estimating the true proportion of defective items in the field. This confidence interval is given by (5). Here, $\hat{p}$ is the proportion of defective items observed in the sample, $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of a standard normal distribution, and n is the sample size.

$\hat{p} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$	(5)

This formula depends on the approximation of the binomial distribution as a normal distribution. This approximation is justified by the central limit theorem, which states that randomly observed variables approach a normal distribution at 30 observations. However, particularly low proportions (close to zero) or particularly high proportions (close to one) may require sample sizes significantly greater than 30 [2]. See (6) to calculate the sample size [9].

$n = [(z_{\alpha/2}\sqrt{(p_0(1-p_0))} + z_{\beta}\sqrt{(p\ (1-p))}) / (p-p_0)\ ]\ \hat{}2$	(6)

The traffic devices being used in Florida are speculated to have a low defect percentage. Thus, the sample size required to construct a confidence interval may be high. Even so, a confidence interval may better answer the question of, "what percent of devices are defective?" than the acceptance sampling plans that have been analyzed thus far. However, the categorization of devices is crucial for developing confidence intervals of similar devices. This issue is dependent on device type, defect type, and vendor.

With this in mind, it is not feasible to construct confidence intervals for every device based on device type, defect type, and vendor. So, criteria are needed to aid in the selection of device. One important criterion is that sampling should be inexpensive whenever possible. Also, the device test should be non-destructive whenever possible. Another criterion may be to choose devices with which previous defects have been a problem.

It is the responsibility of districts to perform field testing. While this responsibility does not fall on the shoulders of TERL personnel, TERL personnel still can provide guidelines for districts to follow. To make sure that these guidelines are followed, they could become a mandatory obligation for passing their QAR.

Intersection Sampling

Sivam Ramalingam (TERL) suggested that the device sampling procedure under development be tweaked to accommodate the resource restrictions that have been discussed. Rather than target the many different types of devices for sampling, intersections themselves should be sampled. After this interview, it was apparent that this maintaining agency already performs preventative maintenance procedures on intersections. Developing a procedure for them to follow would not create substantial extra work for these personnel.

To make this idea work, the first step that needs to be done is for the TERL research team to put together a checklist of what needs to be checked at each intersection. It is understood that intersections come in many shapes and sizes, but in general signaled intersections contain roughly the same types of components. By sampling these components and recording how the components fail, information will be gleaned regarding how they work individually as well as how they perform as part of a system. Once sampled, the maintaining agencies will report their findings back to the TERL, and at that point their work is done.

The next step will be organizing the incoming data. The results should be categorized such that records of which devices from which vendors are performing well and which aren't are easily classified and accessed. Some possible organization schemes may be Excel spreadsheets or Access databases. The organization scheme should lend itself to an efficient statistical analysis tool. Some ideas for the statistical analysis tool are outlined in the monthly progress reports for this project and include acceptance sampling, confidence interval construction, etc. Some new ideas for the statistical tool may be in the field of Reliability Engineering, and further research in such techniques will be performed in the upcoming months. For more information on the idea of intersection sampling, please refer to Sivam Ramalingam's report from this interview.

Reliability Engineering Methods

Reliability, in essence, is the ability of a product, device, system, vehicle, machine, etc. to perform its intended function over a pre-specified period of time. In other words, reliability is quality over time. Statistical methods for quantifying such reliability are available and generally involve the collection and analysis of reliability data from designed experiments in controlled environments, tests on prototypes, close scrutiny of early production units in the field, and methodical, long term tracking of products in the field. Reliability data is traditionally in the form of recorded time of failure for a particular unit; however, other techniques (such as degradation leading to failure) can also be useful. Such data is useful for estimating a unit's hazard function, or failure rate, over time.

There are many reasons for collecting reliability data. Some reasons of particular interest are to assess whether government regulations have been met (or customer specifications), tracing a product throughout its life in the field to determine the reason of failure, identifying methods to improve product reliability, and contrasting components from multiple manufacturers, operating conditions, materials, production periods, etc. Usually, incoming reliability data is not perfect in the sense that exact failure times are unknown. Such data is called censored data, or data that provides bounds on the exact failure time of a product. Statistical models from reliability data typically make use of positive random variable distributions, such as the Weibull, gamma, and lognormal distribution [8].

Censoring

When observing failure times of products, there are frequently constraints that inhibit the ability to record exact failure times. For example, consider a situation in which ten mast arms are installed in an intersection, and after two years, two have them have broken. It is known that the remaining eight have not yet failed, but that they will sometime in the future after two years. Since the data is not complete, it is considered “censored data.” The censoring time of data is simply the time at which the monitoring of unfailed units is discontinued.

There are two general forms of censoring. The first type is referred to as time censoring, or type I censoring. This involves removing unfailed units from a test once a preconceived time is reached in the test. The other type is referred to as failure censoring, or type II censoring. This involves testing until a predetermined number of units fail. This type of censoring can be problematic in that the amount of time for such a test is unknown and could potentially last longer than the investigator wishes to take. However, the statistical properties of type II censoring are less complicated than type I censoring.

In some cases, failures are detected during periodic inspections. Such failure times are known only to be within an upper and lower bound, and are indicative of interval censoring. In other cases failures are detected on the first inspection. Such failures are known to have occurred only before the first observation, and are indicative of left censoring. Furthermore, some failures are known only to have occurred after a certain benchmark in the observation time interval. This is common in the case that some units have not yet failed. All that is known is that they have not failed up to a certain time and therefore their failure time is after that time. Such data is indicative of right censored data [8]..

Introductory Reliability Statistics

Assume that the failure time of a unit, T, is a positive random variable, it can be characterized by a cumulative distribution function (cdf), a probability distribution function (pdf), a survival function, and a hazard function. The x axis for all of these functions is time, t. This interval is separated into m+1 separate observation intervals. For the cdf, the y axis F(t) can be viewed as the proportion of units that will fail before time t. For the pdf, the y axis f(t) represents the relative frequency of failure times as a function of time. The survivability function's y axis S(t) gives the probability of surviving beyond time t. The hazard function's y axis h(t) gives the instantaneous failure rate at a given time t.

Let π_i equal the multnomial probability that a unit will fail in the interval i. It is given in (7). Think of it as the proportion of units that will fail at time t_i less the proportion of units that will fail at time t_{i-1}.

$\Pi_i = \Pr(t_{i-1} < T \le t_i) = F(t_i) - F(t_{i-1})$	(7)

Let p_i equal the conditional probability that a unit will fail in the interval i, given that it was working at the beginning of interval i. It is given in (8). One may think of it as the proportion of failures at time t_i less the proportion of failures at time t_{i-1}, which is then divided by 1 less the proportion of failures at time t_{i-1}.

$p_i = \Pr(t_{i-1} < T \le t_i \mid T > t_{i-1}) = \pi_i / S(t_{i-1})$	(8)

The likelihood function is either equal to or close to the probability of a data set. It changes for different types of censored data. For interval censoring, the likelihood function is given as $L_i(\mathbf{p}) = F(t_i)-F(t_{i-1})$. For left censored observations, the likelihood function is given as $L_i(\mathbf{p}) = F(t_i)$. For right censored data, the likelihood function is given as $L_i(\mathbf{p}) = 1- F(t_i)$.

Nonparametric Estimation

To compute a nonparametric estimate of a cdf, the type of censoring needs to be determined. For interval-censored or right censored data, let n be the original number of units at time 0, let d_i be the number of units that fail during interval I $(t_{i-1}, t_i]$. Thus, the nonparametric estimator of the cdf $F(t_i)$ is given in (9).

$\hat{F}(t_i)$ = # of failures up to time t_i / n	(9)

This estimate is based on the binomial distribution and is the maximum likelihood estimator of F(t). This estimate is descriptive of the sample drawn, and it may also be descriptive of the population from which the sample was taken. However, the estimate of F(t) will differ from the true value of F(t), so a point estimate of F(t) may be misleading. A confidence interval on F(t) quantifies uncertainty due to "sampling error." Confidence intervals are essentially probability statements that the procedure used to construct the interval will contain the true value of interest. A conservative confidence interval on $\hat{F}$ is given in [8], page 50. On the same page, another point wise normal-approximation confidence interval is given. However, for the latter to be adequate, the number of samples multiplied by the estimated F(t) should be between 5 and 10.

Multiple censored data in which failures occur after some units have been censored requires another type of estimation. For a given unit that does not fail during interval i, it may either be censored after the interval ends or it progresses to interval i+1. As long as intervals for different units do not overlap, the following mechanism can be used for estimation. First, let d_i be the same as above and let r_i be the number of units that survive interval i but are right-censored at the end of this interval (t_i). Next, let n_i be equal to n (the total sample size at the first interval) minus the sum of the units that fail over all intervals (dj) minus the sum of all units that are right censored at the ends of various intervals (r_j). The sample proportion failing for an interval i can be used as an estimator of the conditional probability of failing in interval i, given that a unit has made it to interval i. This value is equivalent to the sum of the units that fail over all of the intervals divided by $n_{i.}$. Note that there are separate values for this quanitity for each interval. Using the equation on page 53 of [8] for the survivability function, one can determine an estimate of $F(t_i)$. More types of such point wise confidence intervals are given in Chapter 3 of [8].

Point wise confidence intervals are useful for estimating the cdf of a function at a particular moment in the test time. However, to quantify sampling uncertainty as well as the cdf, it is useful to use simultaneous confidence bands over F(t). An example of large sample simultaneous confidence band is given on page 61 of [8].

Parametric Models for Reliability Data

So far, some of the models used to describe reliability data have been in the form of curves, such as the cdf or the survivability function. However, parametric models can more concisely describe the same data and provide smooth approximations of failure-time distributions. Assume that θ is the single scalar parameter of a given distribution. This value becomes incorporated into various estimates, such as the probability to failure, the "p quantile," the hazard function, the mean life, and the variance of T (page 77). The Var(T) is the averaged squared deviation of T from its mean, and taking the square root of this value yields the standard deviation of T (which is in the same units as T). The variance of T is known as the second moment of T. Other moments exist, but in most reliability applications, quantiles, failure probabilities, and the hazard function are more meaningful than these moments.

Location-scale distributions include most statistical distributions that are widely used such as the exponential, normal, lognormal, Weibull, logistic, and loglogistic. They are convenient because the theory behind them is somewhat simple. T is said to follow an exponential distribution if $T \sim EXP(\theta, \gamma)$. These two parameters can be used to characterize the distribution's cdf, pdf, and hf. The mean of the exponential distribution is $E(T)= \gamma + \theta$ and $Var(T)=\theta^2$ where γ is a location and threshold parameter (that is commonly equal to zero) and θ is a scale parameter. The exponential distribution is commonly used in reliability data; it is characterized by a constant hf. In other words, the propensity of a unit to fail is independent of its age. This is a popular choice for some electronic components assuming they do not have failure causing defects. It would also be useful for physical components that fail due to wear long after their expected life. However, this distribution is usually inappropriate for modeling the life of mechanical components.

Another important distribution is the normal distribution. A random variable Y is said to be normally distributed when $Y \sim N(\mu, \sigma)$. In this case μ is a location parameter (the mean) and σ is a scale parameter (the standard deviation). This distribution is often used as a model for variability, and it has a long history of use due to the central limit theorem. This theorem states that "the sum of a large number of independent identically distributed random quantities has, approximately, a normal distribution [8]. However, this distribution is not used that often in reliability data. It is characterized by a hf that increases rapidly slightly before the median life. It has proved useful for electric filament devices and strength of wire bonds. The normal distribution is more common when the logarithm of the failure time is analyzed (page 82).

If T has a lognormal distribution, this is expressed by $T \sim LOGNOR(\mu,\sigma)$. If this is true, so is the following: $\log(T) \sim N(\mu,\sigma)$. Here, μ is a location parameter and σ is a shape parameter. It is common convention to use natural logarithm with base e. This model is a common model for failures caused by degradation processes with some random component. It is widely used to characterize failure times such as fracture from cracks in metals. The hf for this distribution starts at zero, increases to a certain point, then decreases very gradually toward zero again.

The Weibull distribution depends on two parameters, β and η. In this case, β is a shape parameter and η is a scale parameter. This distribution is useful because of its ability to describe failure distributions with assorted shapes. For example, when β is between zero and one, it has a decreasing hf, but when β is greater than one, it has an increasing hf. So, if T has a Weibull distribution, it is expressed as $T \sim WEIB(\mu, \sigma)$ where $\mu = \log(\eta)$ and $\sigma = 1/\beta$. Based on extreme value theory, the Weibull distribution is suitable for modeling failure times.

The logistic distribution is another location-scale distribution that is similar in shape to the normal distribution, except that it is characterized by longer tails. When Y follows a logistic distribution, this is expressed by $Y \sim LOGIS(\mu, \sigma)$. Due to the similarity to the normal distribution, the loglogistic distribution is used. When Y follows a loglogistic distribution, this expressed by $Y \sim LOGLOGIS(\mu, \sigma)$. The shape of this distribution is similar to that of the lognormal distribution. Other distributions that are sometimes used in reliability applications include the smallest extreme value distribution and the largest extreme value distribution.

Chapter Four: Conclusion

The first goal of this project was to investigate current field sampling/inspection requirements to determine the current level of quality assurance being emphasized in the field. Actions to complete this goal incorporated contacting personnel throughout the state at the district and maintaining agency level, interviewing them, educating them about the importance of quality assurance if necessary, and promoting operations at the TERL. So far, interviews have been conducted in Tallahassee and Bay County, which are two very capable maintaining agencies in District Three. The results of these interviews are very promising in these two areas; there are no major complaints about poor product quality. However, due to the limited number of interviews, it is recommended that more contacts be made throughout the state to assess the level of quality in other areas of Florida. The two maintaining agencies that were interviewed do a excellent job and have the necessary resources to do so. Thus, results from these interviews may be biased and further interviews with maintaining agencies from areas with different population densities and different levels of maturation will provide a more well rounded and accurate portrayal of the way things really are.

Furthermore, these interviews strengthened the communication link between the TERL and the maintaining agencies. One of the benefits of such a communication link is the ability of TERL personnel to receive feedback from end users of APL devices. Valuable recommendations have been gathered from this feedback. For example, end users feel the APL web site is very sufficient and needs no improvement, but a web site dedicated solely to maintaining agencies would be a great way to foster communication between end users and TERL personnel. Such a web site would be a prime medium for posting technical bulletins regarding the problems/issues with APL devices in different parts of the state. It would also be a great opportunity to develop an online forum in which maintaining agencies can converse about the APL and quality related topics.

In addition to interviewing end users, the author audited a vendor throughout the course of this research to further investigate the level of quality APL devices in the state. The audit involved a document review session and a plant tour. The documents strongly indicated that the company was dedicated to an effective QA system. A plant tour reinforced this and demonstrated the QA system in action.

The second goal of this project was to develop a statistically sound sampling method that will provide the level of confidence needed to ensure devices placed in the field are of consistently high quality, operate well within their life-cycle, and are compliant to FDOT's specifications. Several candidate sampling techniques for sampling individual devices have been proposed. These candidate techniques include hypothesis testing, acceptance sampling (such as single, double, and sequential acceptance sampling), Wald's sequential probability ratio test (SPRT), confidence interval construction, and statistical methods for reliability data. However, carrying out these statistically sound methods requires large numbers in terms of sample size. For devices such as hardware, signal brackets, etc., larger sample sizes may be feasible, but for more expensive devices such as traffic controllers or dynamic message signs, large sample sizes are impossible. Because of this, these statistical sampling strategies will not be feasible for field sampling of individual devices.

An alternative to device sampling is intersection sampling. This idea involves examining intersections as whole and recording defects at a system level rather than an individual device level. This practice is already being done by some end users (such as Bay County); the next step would be to relay the information back to TERL personnel in an organized way. Another alternative to field sampling is to sample devices prior to installation. If deployed successfully, such alternatives to device sampling would potentially enable TERL personnel to make use of the aforementioned statistical sampling methods. For this to work, end users would need to follow a set of the following best practices:

- Document tests that are performed prior to installation. Be specific by recording what devices were tested, how many in total were tested, how many passed the test, how many failed the test, and what types of failures were observed.
- Relay the results of the tests to FDOT TERL staff.
- Keep logs at intersections that record installation dates of devices, problems, failures, complaints, and repairs.
- After specified intervals, relay summaries of these logs to the FDOT TERL.
- Make use of the APL web site, device complaint form, and other resources provided by the FDOT TERL. Feel free to make suggestions to improve any of these resources.
- If using UPS and/or a battery backup system, initiate routine checkups to test the charge level of the batteries.
- Inspect intersections as contractors install devices, making sure that the devices being installed meet specifications.

References

[1] Approved Product List of Traffic Control Signals and Signal Devices. http://www3.dot.state.fl.us/trafficcontrolproducts/ May 2007.

[2] Binomial Proportion Confidence Interval. Wikipedia Online Encyclopedia. **http://en.wikipedia.org/wiki/Binomial_proportion_confidence_interval July 2007**.

[3] Bryan, Wayne. Personal Interview. August 7, 2007.

[4] Dobson, Don. Personal Interview. September 25, 2007.

[5] Ekambaram, S.K. The Statistical Basis of Acceptance Sampling: An Introduction for Industrial Executives. Asia Publishing House. 1963.

[6] Lisansky, Michael. Single Acceptance and Double Sampling Plans. The Florida Department of Transportation Research Center.

[7] Martin, A. and Simpson, J. Development of a Statistical Sampling Method to Measure Quality of Traffic Devices in the Sate of Florida. Version 4. 2007.

[8] Meeker, W. Q. Statistical Methods for Reliability Data. John Wiley & and Sons, Inc. New York. 1998.

[9] Montgomery, Douglas C. and Runger, George C. Applied Statistics and Probability for Engineers. John Wiley & Sons, Inc. Third Edition 2003.

[10] Morgan, Jeffrey. Personal Interview. May 15, 2007.

[11] Morse, Carl. Personal Interview. May 16, 2007.

[12] Powell, Randy. Personal Interview. September 25, 2007.

[13] Quality Assurance Evaluation Survey. FDOT Traffic Engineering Research Laboratory. April 2006.

[14] R Development Core Team. R: A Language and Environment for Statistical Computing. http://www.R-project.org 2006.

[15] Simpson, Jim. Quality Assurance Monitoring & Sampling Method Development for ITS and Traffic Control Signal Devices. The Florida Department of Transportation Research Center. December, 2006.

[16] Statistical Research Group, Columbia University. Sampling Inspection: Principles, Procedures, and Tables for Single, Double, and Sequential Sampling in Acceptance Inspection and Quality Control Based on Percent Defective. McGraw-Hill Book Company, Inc. 1948.

[17] Syfrett, Gene. Personal Interview. September 25, 2007.

[18] Wetherill, G. Barrie and Glazebrook, Kevin D. Sequential Methods in Statistics (3^{rd} edition) Chapman and Hall. 1986.

Appendix 1: R Program

```
#Function to compute the logarithmic scoring procedure for Wald's SPRT

#alpha = probability of type 1 error (prop0 is true, but decide for prop1)
#beta = probability of type 2 error (prop1 is true, but decide for prop0)
#r=number of defective observation
#s=number of effective observations
#prop0=  H0 states that true proportion = prop0
#prop1= H1 states that true proportion = prop1

 SPRT1<-function(alpha, beta, r, s, prop0, prop1)
 {
  lowlim<-log10(beta/(1-alpha));
  highlim<-log10((1-beta)/alpha);
  ln<-r*log10(prop1/prop0) + s*log10((1-prop1)/(1-prop0));
  Result<-c(lowlim,ln,highlim);
  if (ln<=lowlim)
  {
   s <- sprintf("Low Limit:  %f",Result[1]);
   print(s);
   s2 <- sprintf("Likelihood Ratio:  %f", Result[2]);
   print(s2);
   s <- paste(s,s2);
   s2 <- sprintf("High Limit:  %f", Result[3]);
   print(s2);
   s <- paste(s,s2);
   s2 <- sprintf("Reject H0, decide true proportion is %f", prop1)
   print(s2);
   s <- paste(s,s2);
  } else if (ln>=highlim)
  {

   s <- sprintf("Low Limit:  %f",Result[1]);
   print(s);
   s2 <- sprintf("Likelihood Ratio:  %f", Result[2]);
   print(s2);
   s <- paste(s,s2);
   s2 <- sprintf("High Limit:  %f", Result[3]);
   print(s2);
   s <- paste(s,s2);
   s2 <- sprintf("Fail to reject H0, decide true proportion is %f", prop0)
   print(s2);
   s <- paste(s,s2);
  } else
```

```
  {

    s <- sprintf("Low Limit:  %f",Result[1]);
    print(s);
    s2 <- sprintf("Likelihood Ratio:  %f", Result[2]);
    print(s2);
    s <- paste(s,s2);
    s2 <- sprintf("High Limit:  %f", Result[3]);
    print(s2);
    s <- paste(s,s2);
    s2 <- sprintf("More sampling required")
    print(s2);
    s <- paste(s,s2);
  }
}
```

Appendix 2: Acronym List

1) APL: Approved product list
2) ASN: Average sample number
3) ITS: Intelligent traffic systems
4) DMS: Dynamic message signs
5) DOT: Department of transportation
6) QA: Quality assurance
7) QAR: Quality assurance review
8) SPRT: Sequential probability ratio test
9) TERL: Traffic engineering and research Lab
10) TMC: Traffic management center