Evaluation of Texas Condition Assessment Program and Recommendations for Improvement

by

Hui Wu, Ph.D. (corresponding author)
Postdoctoral Research Fellow
Phone: (512)232-3019
Email: plutoing@gmail.com

Zhanmin Zhang, Ph.D.
Associate Professor
Phone: (512)471-4534
Email: z.zhang@mail.utexas.edu

and

Abdus Shakur Qazi,
Graduate Research Assistant
Phone: (512)698-2595
E-mail: shakurqazi@yahoo.com

Department of Civil, Architectural & Environmental Engineering
The University of Texas at Austin

<table>
<thead>
<tr>
<th>Text</th>
<th>4582</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table</td>
<td>750</td>
</tr>
<tr>
<td>Figure</td>
<td>1000</td>
</tr>
<tr>
<td>Total</td>
<td>6332</td>
</tr>
</tbody>
</table>

Paper Submitted for Presentation and Publication at the 92nd Annual Meeting of the Transportation Research Board

Paper revised from original submittal.
ABSTRACT

The Texas Department of Transportation (TxDOT) uses the Texas Condition Assessment Program (TxCAP) to measure and compare the overall road maintenance conditions among its 25 districts. TxCAP combines data from its three subsystems: the Pavement Management Information System (PMIS), which scores the condition of pavement; the Texas Maintenance Assessment Program (TxMAP), which evaluates roadside conditions; and the Texas Traffic Assessment Program (TxTAP), which assesses the condition of signs, work zones, railroad crossings, and other traffic elements. The scores for each of the subsystems are based on data of different sample sizes, accuracy, and levels of variations. Therefore, whether the use of the current TxCAP system is an effective and consistent means to measure the TxDOT roadway inventory conditions needs to be evaluated. Statistical analyses were carried out to evaluate the system from two aspects: 1) to determine whether the mechanism employed in TxCAP is effective in measuring the maintenance performance of 25 districts statistically, and 2) to determine whether the difference between the TxCAP scores for any two districts is statistically significant in terms of the insufficient sampling of the subsystems. A case study was carried out using sample data collected for the whole state from the year 2008 to 2010. The case study results show that the differences in scores between two districts are statistically significant for some of the districts and insignificant for some other districts. It is therefore recommended that TxDOT either compare the 25 districts by groups/tiers or increase the sample size of the data being collected if it wants to compare the districts as individual ones.
INTRODUCTION

Civil infrastructures, particularly roads and bridges, are intended to provide service in terms of effectively, efficiently, and safely transporting both people and goods under all weather conditions. In order to achieve this objective, engineers in highway agencies not only need to maintain the infrastructure to ensure its continued service but also need to be able to evaluate the maintenance process and needs of the roadway network. To ensure that the maintenance activities are effective in terms of performance and cost, engineers and decision-makers need tools to conduct performance-based evaluations. Different agencies have developed various scores and indices to evaluate the performance of the roadway network maintenance. The highway system performance not only depends on the individual performance of pavements and bridges but also on the combined interactive “function” of the pavement component, the traffic component (mainly traffic control devices), and the roadside component. Each of these components or subsystems functions differently and has different maintenance requirements. To be able to evaluate the overall performance of the network, highway agencies need a system to comprehensively plan, measure, and manage the highway system. Such an assessment program/system must be able to organize infrastructure inventory, assess condition, set minimum acceptable condition levels, and establish condition targets. Such a system should serve as a valuable tool to facilitate: a) ranking of projects, b) evaluation of the maintenance needs of the highway network, c) distinguishing between the implications of different performance goals, d) performing cost-effectiveness studies using a performance-based budget selection process, and e) communicating to key customers, including the public, the impact of policy and budget decisions on program service and program delivery (1).

In order to address this issue, the Texas Department of Transportation (TxDOT) developed the Texas Condition Assessment Program (TxCAP) to assess the state’s interstate and non-interstate highways in a more comprehensive manner. TxCAP combines data from the three existing systems: Pavement Management Information System (PMIS), Texas Traffic Assessment Program (TxTAP), and Texas Maintenance Assessment Program (TxMAP). As a result, TxCAP eliminates duplication of the three separate scoring systems and provides a simplified scoring system. However, the current TxCAP consists of three subsystems where scores are calculated using data of different sample sizes, accuracy, and levels of variations. Whether the use of TxCAP is an effective and consistent means to measure the TxDOT roadway maintenance conditions raises concerns and needs to be evaluated (2).

In this study, an attempt is made to analyze the source data and evaluate the validity of the current TxCAP system. The rest of the discussions are organized as follows. In Section 2, a thorough literature review of state of the art and the practice in performance-based condition assessment programs is presented. Section 3 elaborates on the current practice by TxDOT and how the calculations are conducted. Section 4 describes the data used in the case study to evaluate the current TxCAP system. Section 5 presents the methodology and the results of the case study. Finally, Section 6 provides the conclusions and recommendations.

LITERATURE REVIEW

Several states have developed different practices and methodologies that integrate infrastructure inventory, condition assessment, minimum acceptable condition levels, and funding decisions. A review of the practices followed by the Florida Department of Transportation (FDOT), the Ohio Department of Transportation (ODOT), the Oregon Department of Transportation (ODOT), TxDOT, and the Washington Department of Transportation (WSDOT) revealed that not all
infrastructure assets were assessed—only major ones were, such as pavements and bridges (1). However, a few agencies had programs either in operational phase or under development to establish condition targets for other high value assets such as drainage structures and traffic control devices. Table 1 summarizes the practices for overall road inventory condition assessment of five state DOTs.

As it can be seen from Table 1, all the five DOTs developed condition assessment procedures for pavements and bridges. Some DOTs also use the latest mapping technologies such as Geographic Information Systems (GIS) and the Global Positioning Satellites (GPS). The Florida DOT, Texas DOT, and Washington SDOT use automated condition assessment vehicles for pavement condition assessments. The Ohio DOT and Oregon DOT, on the other hand, employ a manual approach.

### TABLE 1 Practices for Overall Road Inventory Condition Assessment by State DOTs

<table>
<thead>
<tr>
<th>AASHTO Member Department</th>
<th>Measurement Component</th>
<th>Comprehensive Road Inventory Scoring System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pavements</td>
<td>Roadside Maintenance</td>
</tr>
<tr>
<td>Florida DOT</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ohio DOT</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Oregon DOT</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Texas DOT</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Washington DOT</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

*Culvert inspection under development and inspection by some districts

More specifically, the Florida DOT maintains a Roadways Characteristics Inventory (RCI) and a Maintenance Rating Program (MRP), which is a visual and technical evaluation of actual highway maintenance conditions. The MRP is divided into five categories (elements): Roadway, Roadside, Traffic Services, Drainage, and Vegetation and Aesthetics; and each element is further divided into characteristics. Using a weighting process, characteristic and element scores are converted into the overall MRP Score. Scores for maintenance areas are rolled up into an overall Statewide MRP Score, which is then reported to the Executive Committee as Florida’s “Report Card.”

A similar methodology is followed by the Washington DOT. The Washington DOT currently uses the Maintenance Accountability Process (MAP) for comprehensively planning, measuring, and managing the state highway system. A sub program of this system is the Maintenance on State System (M2) Program, which contains the maintenance functions needed to maintain the highway system. Each of these groups has its associated MAP activities. The groups are Roadway Maintenance & Operations, Drainage Maintenance & Slope Repair, Roadside & Vegetation Management, Bridge & Urban Tunnel Maintenance & Operations, Snow & Ice Control Operations, Traffic Control Maintenance & Operations, and Rest Area Operations. MAP utilizes outcome-based performance measures with a rating scale of A (best) to F (worst) for reporting the level of service provided. The system also allows maintenance activities to be prioritized according to their impact on specific policy objectives.

**THE TEXAS CONDITION ASSESSMENT PROGRAM**

TxCAP combines information from PMIS, TxMAP, and TxTAP. As a result, TxCAP provides a more comprehensive assessment of the interstate and non-interstate highway system (2).
PMIS is an automated system for storing, retrieving, analyzing, and reporting pavement condition information such as distress, ride quality, deflection, and skid resistance data. It can be used to retrieve and analyze pavement information in order to compare maintenance and rehabilitation treatment alternatives, monitor current pavement condition and estimate total pavement needs. The annual PMIS survey currently consists of three separate surveys: a visual evaluation survey, a ride quality survey, and skid resistance measurement (3).

TxMAP evaluates the overall condition levels for the interstate and non-interstate highway systems. TxMAP requires a visual inspection of 23 elements of the highway system in three different components: pavement, roadsides, and traffic operations for each 1-mile segment (in one direction). The entire evaluation procedure requires only two full-time employees who perform the evaluations with assistance from district personnel. TxMAP inspection consists of evaluation of 10 percent of the interstate highway and 5 percent of all other highways in the state highway system (3).

TxTAP evaluates traffic control devices across the state such as signs, work zones, railroad crossings, and other traffic elements. TxTAP is a tool used by TxDOT to evaluate the uniformity, quality, and consistency of traffic control devices placed on the state highway system. Because evaluating every traffic control device statewide is not feasible in terms of available resources, TxTAP scores are based on a relatively small sample of all traffic control devices. TxDOT’s Traffic Operations Division conducts the annual evaluation of the various types of traffic control devices in each of TxDOT’s 25 field districts. Each district review consists of 20–30 randomly selected segments on the state highway system, 5–16 signalized intersections, 3–4 work zones, and 2–6 railroad crossings (3).

EXAMINATION OF THE STRUCTURE OF TxCAP SCORE

The overall inspection score, commonly known as the TxCAP score, is calculated from the weighted average of the three subsystems, i.e., the TxTAP score, the TxMAP score, and the PMIS score. Each of these scores is calculated as the average of the element scores. The elements are scored via field inspections conducted manually by experts and technicians and by automated vehicles. The elements in each of the subsystems are detailed in Figure 1 (3, 4).
The development of TxCAP eliminates duplication of the three separate scoring systems and provides a simplified and concise scoring system, where potential double-counting the effect of any components is prevented. In this sense, the structure of the TxCAP score is valid in terms of the components being considered.

On the other hand, the ratings and descriptions of the numerical grading system used in TxCAP are based on a five-point system. The five-point system then is converted to a percentage by multiplying each rating by 20. The resulting score is then weighted to determine the overall score for each subsystem. Each subsystem’s overall score is then weighted according to the appropriate TxCAP value to obtain a total composite score for the entire roadway system (4). The TxCAP score is thus a weighted average of the scores of its three components. The TxCAP score (for each district) is calculated from its components’ scores using the following formula:

\[ TxCAP = (0.5 \times PMIS) + (0.25 \times TxMAP) + (0.25 \times TxFFP) \]  

In other words, there are a number of weights applied at different levels to the structure of TxCAP score, resulting in a scoring system that can be directly affected by the choice of the weights. For this very reason, the TxCAP scores cannot be used as an absolute measurement of roadway conditions, but can be used as a relative measurement of the maintenance performance of the 25 TxDOT districts. Furthermore, even when the TxCAP score is used as a relative measurement of the maintenance performance, it can be fine-tuned to more accurately reflect the
local conditions of a district, especially considering that Texas covers wide range of climatic conditions in terms of temperature and moisture. For example, in west Texas, vegetation management may not be as important as it is to a district in east Texas, as in general there is not much rainfall in west Texas; consequently, a smaller weight for vegetation management should be used for a district in west Texas than that for a district in east Texas.

**DATA DESCRIPTION**

In order to evaluate the effectiveness and consistency of the current TxCAP system, a case study was carried out using a dataset collected from all 25 districts spanning a period of 3 years: 2008, 2009, and 2010. The dataset consisted of scores of the elements under each of the three subsystems as well as the calculated PMIS, TxTAP, and TxMAP scores for each surveyed section within each district. The TxCAP score for each section was calculated using Equation 1. The respective district scores were obtained by averaging the scores of the sections surveyed in that year. The average of the scores for all districts gave the mean score for the state. The mean and standard deviation of all the four scores (TxCAP, PMIS, TxMAP, and TxTAP), for the 3 years of data used in the case study, are shown in Table 2. The data in Table 2 shows that, of the three subsystems, TxMAP has the least variation and TxTAP has the highest variation in scores. This indicates that condition of the roadside environment is more consistent, whereas the condition of the traffic component has large fluctuations. The analysis was carried out for each year individually and for the 3 years combined together.

<table>
<thead>
<tr>
<th>Year</th>
<th>PMIS Mean</th>
<th>PMIS S.D.</th>
<th>TxTAP Mean</th>
<th>TxTAP S.D.</th>
<th>TxMAP Mean</th>
<th>TxMAP S.D.</th>
<th>TxCAP Mean</th>
<th>TxCAP S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>77.80</td>
<td>11.111</td>
<td>79.48</td>
<td>12.243</td>
<td>81.24</td>
<td>9.004</td>
<td>79.13</td>
<td>8.87</td>
</tr>
<tr>
<td>2009</td>
<td>74.75</td>
<td>9.176</td>
<td>78.55</td>
<td>10.202</td>
<td>78.92</td>
<td>7.189</td>
<td>76.76</td>
<td>7.19</td>
</tr>
<tr>
<td>2010</td>
<td>76.52</td>
<td>8.874</td>
<td>80.10</td>
<td>10.170</td>
<td>79.93</td>
<td>6.844</td>
<td>78.26</td>
<td>6.91</td>
</tr>
<tr>
<td>Combined</td>
<td>76.34</td>
<td>9.838</td>
<td>79.37</td>
<td>10.919</td>
<td>80.02</td>
<td>7.763</td>
<td>77.86</td>
<td>7.98</td>
</tr>
</tbody>
</table>

**EVALUATION OF TxCAP AND RECOMMENDATIONS FOR IMPROVEMENT**

This section covers the data analysis methodology employed in this study. In the first part, the statistical comparison of the scores is discussed, followed by the minimum sample size determination for the subsystems. The results of the data analysis are also presented in this section.

**Comparison of the Performance Scores**

One of the objectives of this study was to determine whether the difference between the TxCAP scores for any two districts is statistically significant or not, and then to determine if all the districts are performing at the same level or not. As discussed above, the TxCAP score is a mean value calculated based on sampling scores from its three sub system for each district. Therefore, the TxCAP score for each district can be considered as a random variable. A null hypothesis of the comparison was constructed as below,

\[ H_0: \bar{X}_{TxCAP,1} = \bar{X}_{TxCAP,2} \]  

and the alternative hypothesis as
where: $\bar{X}_{TxCAP,1}$ = the mean of TxCAP score for district 1, and
$\bar{X}_{TxCAP,2}$ = the mean of TxCAP score for district 2.

The $t$-test is usually used to compare means for two samples from the same population as well as for samples from two different populations (5, 6). A pilot evaluation showed that the variances of the scores in TxCAP subsystems differed among the districts. Therefore, the TxCAP score for each district was considered as an independent random variable with a different variance. A two-tailed, heteroskedastic $t$-test was carried out to compare the mean values, the formulas of which are presented by Equations 4 through 6.

The two possible outputs of the hypothesis test are either “reject $H_0$” or “fail to reject $H_0$.” If $t$-test results in rejecting the null hypothesis, it means that the scores are statistically different and a true difference exists in the scores. If the $t$-test results “fail to reject” the null hypothesis, then it can be concluded that the scores are not statistically different. In such a case the scores of the two districts may be the same or appear different due to variability in measurement (measurement error). The same null and alternate hypothesis was followed in the comparison of the PMIS, TxTAP, and TxMAP scores. The $t$-statistic, combined standard deviation, and the degrees of freedom were calculated using Equations 4 through 6, respectively (5, 6);

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1 - \bar{X}_2}}$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$D. F. = \frac{\left(\frac{s_1^2}{n_1}\right)^2 + \left(\frac{s_2^2}{n_2}\right)^2}{\left(\frac{(s_1^2)^2}{(n_1 - 1)}\right) + \left(\frac{(s_2^2)^2}{(n_2 - 1)}\right)}$$

where: $\bar{X}_1$ = the mean TxCAP score of district 1,
$\bar{X}_2$ = the mean TxCAP score of district 2,
$s_1^2$ = unbiased estimator of the variance of sample data points for district 1,
$s_2^2$ = unbiased estimator of the variance of sample data points for district 2,
$n_1$ = number of sample data points for district 1, and
$n_2$ = number of sample data points for district 2.

Pairwise comparison of the mean for PMIS, TxTAP, TxMAP, and TxCAP scores of the 25 districts were then carried out. The $t$-statistics, calculated using Equation 4, were compared with the critical $t$-statistics at a 95-percent level of confidence. The output results of the $t$-tests are a number of symmetrical matrices for each year of analysis for each of the four performance
scores. Each output is a matrix of “Yes” and “No” results. “Yes” indicates that the difference between the two respective districts is statistically significant, i.e., it can be concluded that the two districts have different levels of performance. Similarly, a “No” response indicates that statistically the two respective districts have the same performance level. In addition, another set of matrices has been created that presents the level of confidence of the scores. In other words, it shows the probability that the scores are different. The second set of matrices can allow engineers to conduct hypothesis tests at many different levels of confidence without having to repeat the calculations each time. Figure 2 shows the \( t \)-test results of the TxCAP scores for the year 2010. It should be noted that more information can be derived from these outputs by grouping the “No” responses as shown in Figure 2. Because each “No” response indicates that the scores of the two respective districts are not statistically different, a group of “No” responses indicates that the scores of all districts within that group are not statistically different. In other words, no true difference exists between them and all members of this set are performing at the same level. Figure 3 presents the same results in a different format, i.e., the level of confidence tables. Instead of conducting a two-tail test at 95 percent, the level of confidence is shown in Figure 3. In other words, Figure 3 shows the probability that the scores are different when conducting a two-tail test. As an example, the cells corresponding to an 80 percent level of confidence, in Figure 3, have been highlighted.

*Cells indicated “No” responses are highlighted.

** FIGURE 2  \( t \)-test results for TxCAP scores for 2010. **
The changes in any of the performance scores between different years or between districts could be due to a number of reasons, including insufficient sample size. A generally recognized statistical rule is that the accuracy of the estimated mean value of a population increases as the number of samples taken from the population increases. The accuracy of the estimate for variability or standard deviation from the mean also increases with the increase in sample size. In other words, the greater the number of samples collected, the higher the confidence level that the mean will be identified with sufficient accuracy and that the variability will be better defined (7).

Many statistical methods are available for determining the sample size, such as the Bootstrap method, the Assume Normal-Pool Variance method, the Noether method, and the Risk-based method. Also, some valuable experience and results can be found in and adopted from related research. For instance, Zhang (7) conducted a thorough research on determining the appropriate minimum testing frequencies for the construction and maintenance of highway infrastructure using the risk-based method, which is one of the most popular and effective methods. The risk is determined by the probability of making a hypothesis testing error, i.e., both Type I and Type II errors, and tolerable error (7, 8). A number of factors affect the determination of a suitable sample size. In Zhang’s research, four factors were considered to determine the optimal sample sizes or appropriate testing frequencies: the variability of the quality characteristic being measured, the risks that a state DOT or a contractor is willing to take, the tolerable errors each party is willing to accept, and the cost of the testing to be performed. A risk-based statistical approach was used with the aim of conducting tests to achieve a certain standard of quality. In this study the first three factors have been considered, but the issue of cost has not been addressed. Since both Type I and Type II errors should be controlled when determining the required sample size, the minimum sample size (considering a two-tailed test) was calculated using the following formula (5, 6, 7):
\[ n = \frac{(Z_{\alpha/2} + Z_\beta)^2 \sigma^2}{e^2} \]  

where,

\( n \) = sample size,
\( Z_{\alpha/2} \) = the \((1 - \alpha/2)^{th}\) percentile of the standard normal distribution; \( \alpha \) being the Type I error,
\( Z_\beta \) = the \((1 - \beta)^{th}\) percentile of the standard normal distribution; \( \beta \) being the Type II error,
\( \sigma \) = standard deviation, and
\( e \) = tolerable error

Equation 7 shows the different factors that affect the determination of minimum sample size. The variance used in calculating the minimum sample size was obtained from the sample data available for the 3 years. The sample sizes were calculated for different combinations of \( \alpha, \beta \) and \( e \). Different sample sizes have correspondingly different risk levels (7). The risk level includes/combines the \( \alpha \) risk, \( \beta \) risk and the risk associated with \( e \). The tolerable error is defined as the maximum difference the decision-maker is willing to conclude that two comparing scores are the same. Different values of tolerance have been mentioned in literature and are determined, in most cases, from experience or by expert’s judgment. This study also attempted to determine a suitable estimate of the tolerable error from the data available. The mean of the differences was used to obtain a reasonable estimate of the tolerable error (\( e \)). A histogram of the differences of each performance score was plotted and the mean of the absolute differences was obtained for TxTAP, TxMAP, and PMIS as being 3.61, 2.92, and 3.61, respectively. Based on these mean values, suitable values of the tolerable error were selected to calculate the required sample size. The values selected for the tolerable error \( e \) are 0.5, 1.0, and 2.0. A total of four different values (0.01, 0.05, 0.1 and 0.2) were selected for both \( \alpha \) and \( \beta \) for each of the three different tolerable errors. Table 3 shows the current data collection practice and the recommended number of samples to be collected in order to conduct tests at a statistical significance level of \( \alpha \) = 0.05, \( \beta \) = 0.05 for the three different tolerable errors. It should be noted that different districts have different sample sizes as detailed in Figure 4, which shows the data collection practice for all 25 districts over the 3 years. Based on Equation 7 and the standard deviation values in Table 2, TxTAP is expected to require the largest number of samples and TxMAP the least to ensure the estimation accuracy at the same risk level. This expectation is verified by the results in Table 3. Thus more data points are required to carry out tests at the same risk level. For the purpose of establishing a valid TxCAP, data for all three subsystems is required for all pavement sections under consideration (being sampled). Therefore it is recommended that the number of data samples collected should match the largest minimum sample size (of the three subsystems) for a chosen risk level. Although PMIS data is collected for the state/wild highways, insufficient data is collected for TxMAP and TxTAP (9).

Two methods were considered to improve the data collection without incurring a significant increase in costs. One is to take data from \( \frac{1}{2} \)-mile segments instead of the current 1-mile segments, i.e., by dividing the current sample into two samples. However, this method does not increase the statistical significance of the data as the data becomes non-random, violating the key assumption for the sampling process. The statistical significance can be increased only if all
sections (1-mile or ½-mile) are randomly selected. Another option is to aggregate the data for 2 years to create a larger sample. This process will definitely increase the sample size, but care must be taken to ensure the time periods for comparison do not overlap.

### TABLE 3 Data Collection Practice for the Three Performance Scores

<table>
<thead>
<tr>
<th>District</th>
<th>PMIS</th>
<th>TxTAP</th>
<th>TxMAP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Min</td>
<td>Max</td>
<td>Avg.</td>
</tr>
<tr>
<td></td>
<td>2008</td>
<td>2009</td>
<td>2010</td>
</tr>
<tr>
<td>PMIS</td>
<td>99</td>
<td>260</td>
<td>154</td>
</tr>
<tr>
<td>TxTAP</td>
<td>99</td>
<td>258</td>
<td>160</td>
</tr>
<tr>
<td>TxMAP</td>
<td>103</td>
<td>257</td>
<td>157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>District</th>
<th>Min</th>
<th>Max</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2008</td>
<td>2009</td>
<td>2010</td>
</tr>
<tr>
<td>PMIS</td>
<td>262</td>
<td>1047</td>
<td>4190</td>
</tr>
<tr>
<td>TxTAP</td>
<td>323</td>
<td>1290</td>
<td>5161</td>
</tr>
<tr>
<td>TxMAP</td>
<td>163</td>
<td>652</td>
<td>2609</td>
</tr>
</tbody>
</table>

### FIGURE 4 Data collection practice and minimum sample size for $\alpha = 0.05$, $\beta = 0.05$, and $e = 2$. 
**CONCLUSION**

The objective of this study is to evaluate whether the use of the current TxCAP system is an effective and consistent means to measure the TxDOT roadway maintenance performance of TxDOT districts, and develop recommendations according to the analysis results. To achieve this objective, the source data of each subsystem of TxCAP was thoroughly analyzed, and the validity of the current TxCAP system developed as a combination of three subsystems is studied. Based on a careful examination of the structure of TxCAP scores, it was clear that the TxCAP scores cannot be used as an absolute measurement of roadway conditions, but can be used as a relative measurement of the maintenance performance of the 25 TxDOT districts. Furthermore, it is recommended that, even when the TxCAP scores are used as a relative measurement of the maintenance performance of the TxDOT districts, it should be enhanced to more accurately reflect the local conditions of the districts through adjusting the weights that are used at various levels of the TxCAP structure.

The statistical analyses were carried out in two steps on a dataset for all 25 districts spanning over a period of years. In the first step, two-sample comparison was conducted to test the null hypothesis: the mean TxCAP scores of two districts are equal. A two-tail, heteroskedastic, t-test with a 95 percent level of confidence was carried out on the sample data to test the hypothesis. Each district was treated as a separate population with a different size and variance. The results, presented as matrices, showed which pairs of district scores are statistically different and which are not. The results can be arranged to form groups or tiers of statistically similar districts. Within each group, statistically no difference exists between the districts, i.e., the maintenance performance of those districts are statistically the same. In addition, a matrix of the level of confidence for TxCAP was created. This matrix gives the probability that scores (of any two districts) are different. The test was also conducted on the subsystem scores.

In the second part of the study, a risk-based statistical procedure was selected for determining the minimum sample size (survey sites) for each of the subsystem. The sample size determination is associated with variability, Type I error, Type II error, and tolerable error. The analysis results show that the three subsystems have different variances, and therefore each subsystem requires a different minimum sample size, for a specific risk level. However to establish a valid and effective TxCAP system, all three subsystem scores are required for all sampled sections. This analysis, therefore, recommends that if TxDOT wants to achieve a higher and uniform risk level for TxCAP, the amount of sample data being collected for TxTAP and TxMAP should be increased, as the PMIS data is already being collected for the entire network of the state highways; however, as an alternative, TxDOT can maintain the current level of data collection effort, but compare the districts by groups/tiers, where the differences in scores for the districts within the same group/tier are statistically insignificant.
REFERENCES


7. Zhang, Z., Fowler, D. W., Hudson, W. R., and Ceballos, R. Development of a methodology to determine the appropriate minimum testing frequency for the construction and maintenance of highway infrastructure, Report number 0-7181-1, Center for Transportation Research, University of Texas at Austin, October 2001.


9. Zhang, Z., Machemehl, R. B., Pavement Related Databases in TxDOT, Report number 0-4186-3, Center for Transportation Research, University of Texas at Austin and Texas Transportation Institute, The Texas A&M University System, August 2004