TRC0905

Evaluation of Aggregate Durability Performance Test Procedures

Stacy G. Williams, Joshua B. Cunningham

Final Report

2012

Evaluation of Aggregate Durability Performance Test Procedures

Final Report

TRC-0905

by

Stacy G. Williams, Ph.D., P.E.
Director, CTTP
Research Associate Professor
Department of Civil Engineering
University of Arkansas

Joshua B. Cunningham, M.S.
Department of Civil Engineering
University of Arkansas

April 26, 2012

1. Introduction

Aggregate properties play a major role in the long-term performance of pavements. An aggregate's quality depends largely on its ability to resist two things: freeze/thaw cycles and physical degradation. The ability to withstand both of these distresses will significantly extend the life of a pavement. As the abundance of high quality aggregates diminishes, tests to evaluate questionable aggregates become more important.

Currently, the Sodium Sulfate Soundness test, AASHTO T 104, is the primary indicator of aggregate soundness used by the Arkansas State Highway and Transportation Department (AHTD). The problem is that this test does not always relate well to actual pavement performance. *(Janoo and Korhonen, 1999; Cuelho, et.al. 2007; Meininger, 2002)* It also tends to be a difficult and time consuming test that yields poor precision because it is highly sensitive to minor differences in procedure and equipment. *(Bloem, 1966; AASHTO, 2009)*

In order to improve the selection of materials, new test methods should be examined. These new tests should include not only aggregate soundness tests, but also tests that are performed on paving mixtures containing the aggregate sources. If a new test method could be established that yields better precision and relates well to laboratory testing of hot mix asphalt (HMA) and Portland cement concrete (PCC), then the life and quality of pavements could be improved.

2. Problem Statement

When soundness tests on limestone and dolomite aggregates fail, limestone and dolomite aggregate suppliers contend that the soundness results do not provide good indications of the durability performance of the limestone and dolomite aggregate. To support the claim, the aggregate suppliers use a research project of limited scope that was performed to evaluate the effectiveness of soundness testing of limestone aggregates as a durability performance indicator. A search of Department Standard Specifications uncovered a copy of the March 1, 1940 Standard Specifications for Road and Bridge Construction. The 1940 Standard Specification required soundness testing of aggregates for Portland cement concrete and asphalt concrete hot mix. Currently the Department specifies AASHTO T 104 Sodium Sulfate Soundness for aggregates. Soundness testing is used as an indicator of the aggregate's durability. For most of the aggregates that are currently used, Sodium Sulfate Soundness testing does not result in a dispute over the durability of the aggregate; however, for some limestone and dolomite aggregates, there is disagreement as to whether the Soundness testing results accurately reflect the field performance, or durability, of the aggregate. Resolution of the debate is necessary to insure that durable limestone and dolomite aggregates are not being disqualified for use on Department projects.

3. Background

Aggregate durability is a characteristic that is critical to the quality of pavements. It is a term that generally describes the resistance of the aggregate to environmental, physical, and cyclical loading conditions, and is affected by temperature, load, moisture, chemical exposure, and freeze/thaw cycles. *(Barksdale, 1991; Williamson, et.al., 2007)* Aggregates with poor durability tend to experience particle breakdown, which leads to gradation changes and serious pavement performance issues. Aggregate durability is a term often used to incorporate the concepts of both soundness and toughness. More accurately described, aggregate *soundness* refers to the aggregate's ability to withstand cyclical environmental distress, while aggregate *toughness* refers to its ability to withstand physical distresses experienced during manufacture, production, transportation, and construction. The methods shown in Table 1 are commonly used to describe the durability of an aggregate source. Methods currently specified by AHTD are noted.

Table 1: Aggregate Soundness and Toughness Tests

Test Method	AASHTO Designation	ASTM Designation	Methodology	Currently Specified by AHTD
L.A. Abrasion	T 96	C 353 / C 131	Abrasion (dry)	X
Sodium Sulfate Soundness	T 104	C 88	Simulated Freeze-thaw	X
Magnesium Sulfate Soundness	T 104	C 88	Simulated Freeze-thaw	
Micro-Deval Durability	T 327	D 6928	Abrasion (wet)	
Aggregate Freeze Thaw	T 103	C 666	Accelerated Freeze-thaw	
Durability Index	T 210	D 3744	Abrasion (wet)	

As indicated in the table, the AHTD currently specifies the Los Angeles (L.A.) Abrasion test to assess toughness and the Sodium Sulfate Soundness test to measure soundness. However, the sodium sulfate test has been found to yield low precision and does not accurately predict an aggregate's performance in pavements. *(Janoo and Korhonen, 1999; Cuelho, et.al., 2007; Meininger, 2002)* For this reason, other soundness tests have been explored to determine whether a different test can better predict an aggregate's performance.

L.A. Abrasion Test

The L.A. Abrasion test is a nationally recognized method for determining the quality of coarse aggregate. In this method, a specifically graded aggregate sample is placed in a revolving drum with steel charges, and rotated for 500 revolutions at a rate of 30 to 33 revolutions per minute. *(AASHTO, 2009)* By comparing the original and resulting gradations of aggregate, a percent loss is calculated. The lower the percent loss, the greater the aggregate's resistance to breakdown caused by impact and abrasion. AHTD currently specifies a maximum of 35 percent loss for aggregates used in HMA pavements, and a maximum of 40% loss for aggregates used in PCC pavements. *(AHTD, 2003)* The L.A. Abrasion machine is shown in Figure 1.

Figure 1. L.A. Abrasion Machine

Sodium Sulfate Soundness

To determine an aggregate's resistance to degradation caused by freezing and thawing, AHTD currently specifies the sodium sulfate soundness test. During this test, aggregates are tested "to determine their resistance to disintegration by saturated solution of sodium sulfate." *(AASHTO, 2009)* This is accomplished by subjecting a specifically graded aggregate sample to repeated cycles of soaking and drying the aggregates in a sodium sulfate solution. During the soaking period, the salt solution enters the aggregate pores. Next, the aggregate sample is oven dried. As the salt solution is dried from the sample, the salt is dehydrated and precipitated in the permeable void spaces within the aggregate. During this phase of conditioning, thawing is simulated. During the next soaking phase, the salts are rehydrated, creating internal expansive forces within the aggregate pores, which simulates the expansion of water during freezing. A series of cycles (usually five) emulates the cumulative effects of repetitive freeze/thaw cycles. At the end of the test, the aggregate grading is analyzed to determine the percent loss of the aggregate sample. Typical limits on percent loss are 12 percent for coarse aggregate and 15 percent for fine aggregate. In Arkansas, aggregate soundness is governed by the sodium sulfate soundness test. Coarse aggregates used in PCC pavements are limited to a maximum loss of 12 percent after five cycles. Likewise, aggregates used in HMA pavements are also limited to 12 percent loss after five cycles. *(AHTD, 2003; AHTD 2009)* The equipment used in this method is shown in Figure 2.

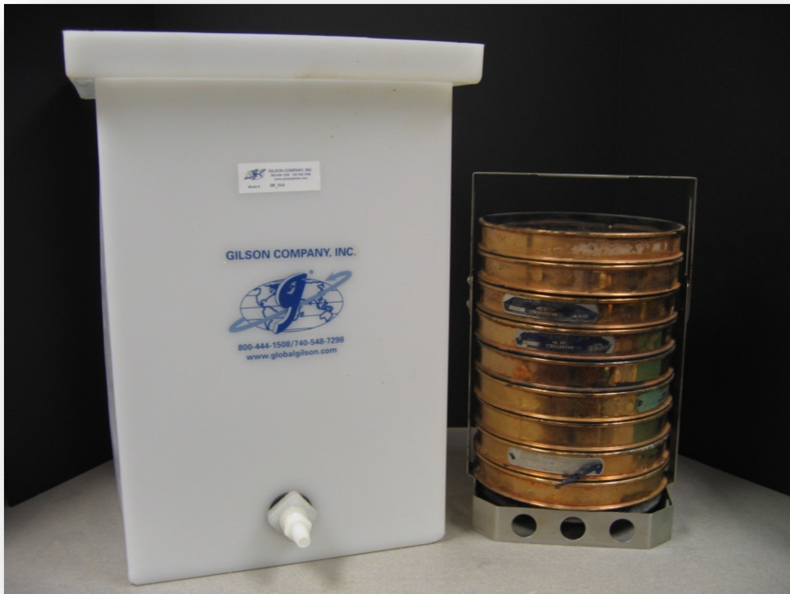

Figure 2. Sulfate Soundness Equipment

The greatest advantage of the sodium sulfate soundness test is that it is fairly common to the pavement industry, and is recognized as a standard test method for aggregate durability. The greatest disadvantage is that test results by this method are not reported to have a strong correlation with actual pavement performance. *(Williamson, et.al., 2007; Wu, et.al., 1998; Cuelho, 2007)* In addition, the method is relatively expensive and time consuming, and has poor precision. The coefficient of variation published for the multilaboratory difference between two tests (D2S%) is 116 percent of the average test result. In addition, a statement is included in the test method that "This test method furnishes information helpful in judging the soundness of aggregates subject to weathering action, particularly when adequate information is not available from service records of the material exposed to actual weathering conditions. . . care must be exercised in fixing proper limits in any specifications that may include requirements for these tests." *(AASHTO, 2009)* In other words, the method provides and indication of soundness, but may not provide an accurate account of the anticipated field performance. Additionally, agencies using this method for acceptance have been known to accept unsound aggregates, while rejecting sound aggregates. *(Bloem, 1966)* For these reasons, it has been suggested that agencies may use the sodium sulfate soundness test to accept aggregates, but not as a single rejection test.

Magnesium Sulfate Soundness

The magnesium sulfate soundness test, also described in AASHTO T 104, uses the same principles as the sodium sulfate method, but uses a different salt to simulate the weathering conditions. In general, the two salts do not provide comparable test results, such that the magnesium sulfate solution creates a greater amount of aggregate breakdown than the sodium sulfate solution. Typical specifications require that the percent loss by magnesium sulfate method be limited to 18 percent for coarse aggregate and 20 percent for fine aggregate. *(Barksdale, 1991)*

The magnesium sulfate alternative is reported to provide greater precision than the sodium sulfate salt, however both are still considered poor. *(Meininger, 2002; AASHTO, 2009)* The greatest disadvantage of this method is, as stated for the sodium sulfate method, that historical field performance of a given aggregate is said to provide more valuable information than the results of this test method. *(AASHTO, 2009)*

Aggregate Freeze-Thaw

The standard method of test for Soundness of Aggregates by Freezing and Thawing, outlined in AASHTO T 103, determines the resistance of an aggregate to disintegration by freezing and thawing by simulating the cumulative effects of weathering. *(AASHTO, 2009)* In this method, an aggregate sample is fractionated and each size fraction is placed in a sample container. The samples may be conditioned by either 1) total immersion in a 0.3 percent NaCl and water solution or 0.5 percent Methyl Alcohol and water solution, 2) partial immersion in a 0.5 percent ethyl alcohol and water solution, or 3) partial immersion in water. After allowing the samples to soak in the chosen solution at room temperature for 24 hours, the samples are cooled to -9°F. This temperature is held for at least thirty minutes, then raised to 70°F and held for thirty minutes, constituting one cycle. This process is repeated for a designated number of cycles (often 50), after which a percent loss is determined. This test, like the sodium sulfate soundness test, is a lengthy process and can take two weeks or more to complete. Also like AASHTO T 104, this method describes cautions that field performance data may be more valuable than test results by AASHTO T 103. *(AASHTO, 2009)*

Some researchers feel that the rapid freezing and thawing creates an unrealistic environmental condition. Field measurements have shown that concrete rarely cools faster than 5°F per hour *(Powers, et.al., 1955)*, and concrete specimens in Ontario, Canada have been shown to rarely experience a cooling rate over 2°C per hour. *(Nokken, et.al., 2004)* It has been found that cooling rate, solution strength, and minimum temperature all affect the percent loss values obtained by this method. *(Hooton and Rogers, 1989)* The precision of this test is also highly affected by the relationship of pore characteristics and aggregate size. The movement of water out of the aggregate, and hence, the durability of the aggregate, is governed by the pore size, porosity, and the aggregate size. *(Powers, et.al., 1955, Verbeck and Landgren, 1960)* Aggregates with larger pores are typically sounder because they have difficulty remaining saturated. Aggregates with finer pores and larger absorption capacities tend to have a higher risk of breakdown. *(Stark, 1976)* However, if the pavement section is prone to retaining water, large-pore aggregates may, in fact, remain saturated.

A similar test is the Canadian Freeze-Thaw test, which was developed by the University of Windsor and the Ontario Ministry of Transportation. The difference between this test and AASHTO T 103 is that the 3 percent NaCl solution is used to simulate the effects of deicing salts. A study by Senior and Rogers *(1991)* indicated that the Canadian freeze-thaw test better represented soundness characteristics than the magnesium sulfate soundness test for asphalt concrete. *(Wu, et.al., 1998)*

Micro-Deval

An increasingly popular test known as the "Resistance of Coarse Aggregate to Degradation by Abrasion in the Micro-Deval Apparatus" is described in AASHTO T 327. This test was developed by the French in the 1960's. *(AASHTO, 2009; Senior and Rogers, 1991)* During this test, a specifically-graded and soaked sample is placed in a mill jar with 20 ± 5°C water and 5 kilograms of steel balls, each 5 mm in diameter. The sample, water, and balls are then revolved at 100 ± 5 rpm for 12,000 ± 100 revolutions. Afterwards the sample is washed and oven dried, and the amount passing the No. 16 sieve is calculated as percent loss. The Micro-Deval device is shown in Figure 3.

Figure 3. Micro-Deval Device

This test is commonly referenced as a toughness test because aggregates are treated in a manner similar to that of the L.A. Abrasion test. However, the aggregates in the Micro-Deval device are tested while wet, and those in the L.A. Abrasion method are tested while dry. The Micro-Deval has been compared to many soundness tests and has been found to have some correlation to the magnesium sulfate soundness test. *(Wu, et.al., 1998)* However, others have reported no relationship between the Micro-Deval test and the L.A. Abrasion or sodium sulfate soundness test. *(Cooley and James, 2004)* The Micro-

Deval method of scouring in the presence of water is believed to be a more accurate representation of the degrading forces applied to aggregates during construction, and describes primarily the resistance of the aggregate to physical degradation. Some believe that because this test is performed with water, it may provide some indication of the aggregate's resistance to weathering. The test has also been referred to as more conservative than the L.A. Abrasion and sodium sulfate soundness tests, meaning that if an aggregate meets the criteria for the Micro-Deval, it will likely also meet the criteria for the other tests. *(Cuelho, et.al., 2008)* It has also demonstrated a better representation of field performance than that of the L.A. Abrasion test for granular bases used PCC pavement construction. *(Senior and Rogers, 1991)* Several studies have reported good precision with the Micro-Deval test and have recommended it as a replacement for the sodium sulfate test.

Aggregate Durability Index

Aggregate Durability Index, described in AASHTO T 210, is also used to determine the toughness of aggregates. The durability index represents the ability of an aggregate to resist the production of "detrimental claylike fines when subjected to prescribed mechanical methods of degradation." *(AASHTO, 2009)* This test was formulated to permit prequalification of aggregates used during the construction of transportation facilities. The test involves washing a sample in a mechanical washing vessel. Afterwards, the fines are collected and mixed with a calcium chloride solution and placed in a cylinder. The height of the sediment is then used to calculate the durability index. The time required to perform this test is shorter than the sulfate soundness test, and even though it is primarily a measure of mechanical degradation, it has also been considered as a replacement for the sodium sulfate soundness test. *(Hamilton, et.al., 1971)*

Aggregate Performance

Since freeze-thaw cycles can be detrimental to pavement performance, aggregates need to be able to withstand the location's climatic changes. In northern states, winter often consists of a continuous cold period and a single (though lengthy) freeze period followed by a "spring thaw", resulting in very few freeze-thaw cycles. Arkansas does not historically experience significant periods of freezing temperatures capable of affecting subgrade soils, but does often experience rapid weather changes generating a large number of short freeze-thaw cycles that can significantly affect the pavement's surface. If the aggregates in the upper portions of the pavement structure are not sound enough to resist these temperature swings, pop-outs or raveling of the pavement's surface can result.

HMA Pavement Distresses

When HMA pavements contain aggregates of poor durability, repetitive freeze/thaw cycles tend to break down the aggregate particles, thereby weakening the aggregate/asphalt bond. When this bond is broken, the pavement becomes susceptible to stripping failures. Stripping, or moisture damage, often begins as a physical breakdown at the bottom of the HMA layer, leading to a loss of support and permanent deformation. Aggregate particles may also loosen from the surface, leading to surface raveling, or a pitted and "pock-marked" appearance.

Stripping is defined as "the progressive functional deterioration of a pavement mixture by loss of the adhesive bond between the asphalt cement and the aggregate surface and/or loss of the cohesive resistance within the asphalt cement principally from the action of water." *(Kiggundu and Roberts, 1988)*

While many studies have been performed to determine the cause of stripping, there is no single soundness test that has been proven to accurately predict stripping. Since the occurrence of stripping continues, it is implied that the based causes of stripping are not fully understood. Some existing theories state that stripping is caused by detachment, displacement, spontaneous emulsification, pore pressure, film rupture, and hydraulic scouring. Explanations for stripping also include mechanical interlock, chemical reaction, molecular orientation, and interracial phenomenon. *(Kiggundu and Roberts, 1988)*

The only factor that is widely recognized to cause stripping is water. Water penetrates the asphalt binder causing stripping. If the infiltration of water can be stopped, an improvement to pavement health and durability would result, mainly because stripping can lead to decreased structural support, rutting, shoving, raveling, and cracking. *(Kiggundu and Roberts, 1988)*

HMA Performance Testing

Many different moisture damage tests have been developed over the years for HMA. Some tests range from simply boiling a specimen to subjecting it to a wheel tracking test. However, the modified Lottman test, AASHTO T 283, is generally specified for Superpave mix designs.

<u>Boiling Test</u>

One of the simplest tests is ASTM D 3625, known as the Boiling Water Test. During this test, loose HMA mix is simply added to boiling water. After a specified period of time, usually 10 minutes, the mix is removed from the water for visual inspection. An acceptable test requires the coated aggregate to retain more than 95 percent of its original binder. Though the test is simple and can be performed quickly, results are subjective, no strength value is calculated, and stripping of fine aggregate is difficult to determine. This method is not recommended for use as a single pass/fail test. *(Williams, 2001)*

<u>Lottman Test</u>

Developed under NCHRP 246, the Lottman Test requires nine samples compacted to expected field air void content. The samples are then divided into groups of three. The first group is the unconditioned control group. The second is vacuum saturated with water for 30 minutes to represent pavement performance after four years, and the third group is vacuum-saturated and subjected to a freeze-thaw cycle intended to represent performance at 4 to 12 years. A split tensile strength test is then run on each sample to determine a ratio of the indirect tensile strength of the conditioned samples to the unconditioned samples. A minimum required ratio of 0.70 is commonly used.

<u>AASHTO T 283</u>

Probably the most commonly used test is a modified version of the Lottman test, known as the Resistance of Compacted Hot Mix Asphalt to Moisture-Induced Damage Test. Described in AASHTO T 283 and shown in Figure 4, this test measures the change in diametral tensile strength of conditioned and unconditioned specimens, where the conditioning process includes vacuum saturation and an optional freeze-thaw cycle. The results of this test can be used to predict long-term stripping susceptibility, and can also be used to assess liquid anti-stripping additives.

Unlike the Lottman test, the Modified Lottman test only involves two subsets of gyratory compacted specimens, with each set consisting of three specimens. One subset is tested for indirect tensile strength in a dry condition, while the other subset is vacuum saturated to 70 to 80 percent and subjected to a freeze cycle followed by a warm water soaking cycle before testing for indirect tensile strength. A retained tensile strength ratio (TSR) of 0.7 to 0.85 is recommended as passing for this test, with 0.80 being the most commonly specified value. *(Williams, 2001)*

Figure 4. Modified Lottman Testing by AASHTO T 283

While AASHTO T 283 is the most commonly used test for determining moisture damage in HMA, highway agencies have reported problems with the test. The most significant shortcoming is that the test does not always accurately predict moisture sensitivity in the field. Samples that yield a high TSR may perform poorly, while samples with a low TSR may perform adequately. Conditioning during the test has also been a factor of concern that affects precision. Results from a NCHRP study showed that samples saturated to 55 percent have displayed significantly different results than similar specimens saturated to 80 percent. *(Azari, 2010)* Poor overall precision, large standard deviations between samples, and poor repeatability between laboratories have also been demonstrated. *(Azari, 2011)*

<u>Wheel Tracking Tests</u>

An increasingly popular way of determining a pavement's moisture sensitivity in the laboratory is through a loaded wheel test (LWT). LWTs have the ability to determine rutting potential as well as stripping potential if a sample is tested in the wet condition. A LWT consists of a loaded moving wheel that travels along a sample's surface, causing depressions. Rut depths are recorded and can be used to provide relative performance comparisons of various mixtures. A LWT is beneficial because it is relatively inexpensive and easy to operate. It can be used in the performance ranking of HMA mix designs, as well as a pass/fail criterion for mix design specifications. However, no value used in mechanistic-empirical design models is determined from a LWT.

The Hamburg Wheel-Tracking Device (HWTD), was developed in the 1970's be Esso A.G. of Hamburg, Germany and is fashioned after a British device that had a rubber tire. Originally named the Esso Wheel-Tracking Device, the City of Hamburg finalized the test method, establishing pass/fail criteria for HMA mixes. Originally used for rutting susceptibility, the test ran for 9,540 wheel passes with a water temperature of either 40°C or 50°C. Later, the number of wheel passes was increased to 19,200 where it was found that samples often started showing the effects of moisture damage after 10,000 passes. *(Williams, 2001; FHWA, 2010)* The German specifications require that a mixture display no more than a rut depth of 4 mm (0.16 in) after 20,000 wheel passes. Specifications in the United States are typically not as harsh. The state of Colorado uses a limiting rut depth of 10 mm (0.4 in) after 20,000 cycles, while the state of Texas allows a maximum rut depth of 12.5mm (0.5 in) after 20,000 cycles. *(Williams, 2001; Wu, et.al., 1998)* The Hamburg LWT is described in AASHTO T 324.

The University of Arkansas developed a device similar to the Hamburg, known as the Evaluator of Rutting and Stripping in Asphalt (ERSA). ERSA, shown in Figure 5, measures the vertical deformation at 40 locations along an HMA specimen and records the rut depths every 100 cycles. Benefits of ERSA are that it can operate simultaneously under wet or dry conditions, and that the air and water temperatures can be adjusted. Typical ERSA output data is shown in Figure 6. This data defines a number of specimen characteristics, including rut depth, rutting slope, stripping slope, and stripping inflection point. A typical sample will experience some initial consolidation, then rut at a relatively constant rate, known as the rutting slope. If the specimen is susceptible to moisture damage, the rate of deformation will increase, generating a stripping slope. The intersection of the rutting and stripping slopes is the stripping inflection point, which defines the point at which moisture sensitivity began to dominate specimen deterioration. *(Hall and Williams, 1998)*

Figure 5. The Evaluator of Rutting and Stripping in Asphalt (ERSA)

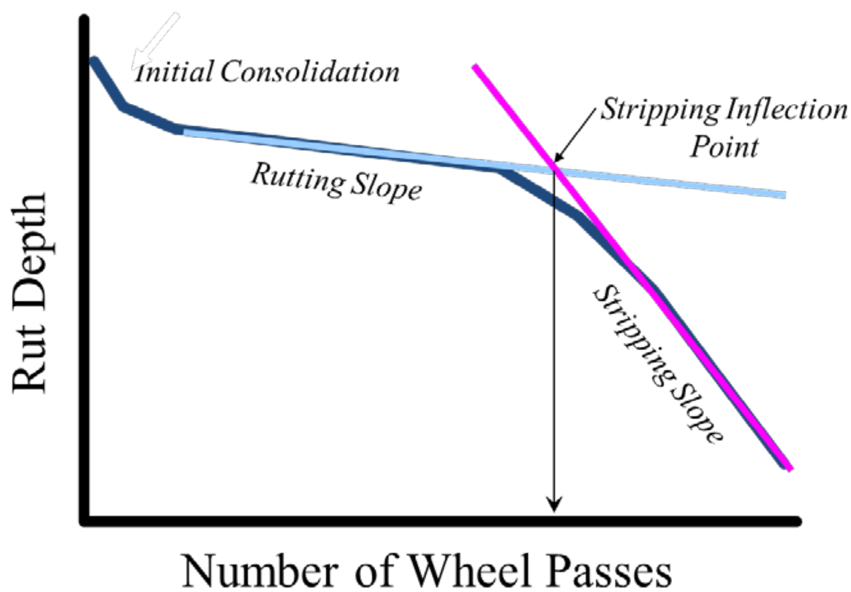

Figure 6. Typical ERSA Output Data Graph

Cantabro Loss

The Texas Contabro Loss test, described in TxDOT Method TEX-245-F, has been used as a relative performance identifier for HMA mixtures. In this method, a compacted asphalt specimen is placed in the Los Angeles Abrasion machine, and tumbled at a speed of 30 to 33 revolutions perminute for 300 revolutions. After tumbling, any loose material that has broken off of the test specimen is discarded, and the weight of the remining portion is measured. The percent loss is calculated from the original sample weight and the weight after tumbling. *(TxDOT, 2005)*

PCC Pavement Distresses

When poor quality aggregates are used, premature pavement failures can occur as a result of freeze-thaw cycles. This is true for saturated concrete pavements that are subjected to freezing conditions because as the water in the pavement layer freezes, it expands causing pressure within the pores of the concrete. The concrete will then rupture isf the pressure exerted is stronger than the tensile strength of the cement paste. After multiple cycles, the concrete breaks down resulting in reduced strength, durability cracking (i.e., D-cracking), map cracking, pitting, and popouts. *(WSDOT, 2010)*

Both D-cracking and map cracking occur when coarse aggregates break down from the expansion forces created from water freezing in the pores of the aggregate. D-cracking appears in the form of a crescent-shaped hairline cracking pattern near joints and pavement edges. Map cracking has a variable pattern, and appears only at the surface of the pavement. Pitting and popouts, shown in Figure 2-4, leave holes at the pavement surface and are a result of poor aggregate freeze-thaw resistance. The holes are typically 25 to 100 mm in length and 13 to 50 mm in depth. *(FHWA, 2003)*

PCC Performance Testing

For PCC pavements, a common test method used to examine how a pavement will perform is ASTM 666, 'Resistance of Concrete to Rapid Freezing and Thawing'. For this test, concrete beams are subjected to freeze/thaw cycles in a freeze-thaw chamber, and the resonant frequency is determined according to ASTM 215 after various numbers of freeze/thaw cycles. The durability factor of the concrete specimen is defined as the ratio of the resonant frequency at a given number of cycles to the resonsant frequency at zero cycles. This factor is calculated regularly until the specimen reaches a total of 300 cycles, or until the beam has lost 60% of its original frequency. Visual inspections of beam deterioration are also noted. *(ASTM, 2010)* The test setup for determining frequency is shown in Figure 7.

Figure 7. Measurement of Resonant Frequency

4. Literature Review

Limestone and dolomite aggregates are prevalent in the northern portion of Arkansas. These sources are shown in pink and brown in the northwest and north central areas of the state, as displayed in Figure 8. Historically, these aggregates have been believed to be susceptible to breakdown due to environmental freeze-thaw cycling, leading to accelerated pavement distress.

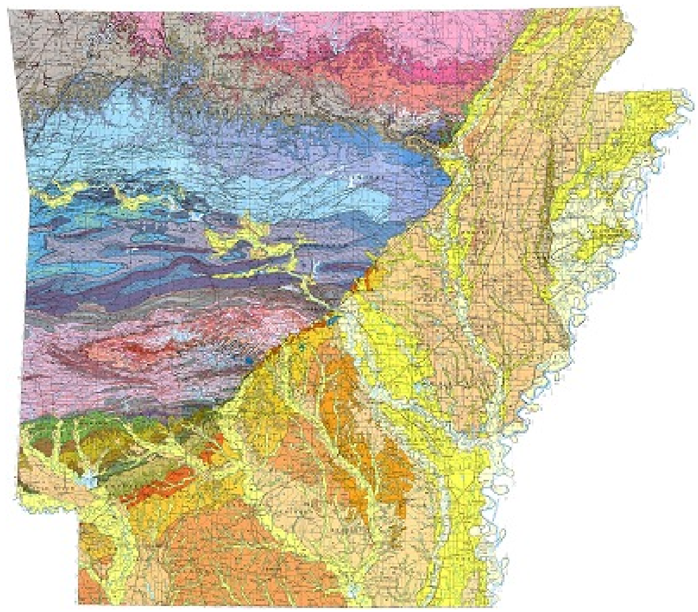

Figure 8. Arkansas Geologic Map *(Geology.about.com, 2009)*

Previous research has been done to assess the applicability of the sodium sulfate soundness test for northern Arkansas dolomite aggregates used in construction *(Kline, et.al., 2004).* In this study, dolomites from the Cotter Dolomite formation were obtained from the Carroll County Stone Company Quarry near Berryville, Arkansas. These aggregates were tested for soundness according to the sodium sulfate soundness methods, and were also analyzed by the insoluble residue, water absorption, and x-ray diffraction analysis methods. Very few significant correlations were developed for the soundness performance of the aggregates, and it was acknowledged that in theory, the primary feature relating to freeze-thaw degradation is the structure of the pore system. Even though very few of the physical features of the aggregates were able to relate to soundness test results, there was a loose correlation between the sodium sulfate soundness test method and performance. Overall, however, it was recommended that the sodium sulfate method be abandoned, or that the specification limits for qualifying aggregates be loosened significantly. This recommendation was supported by Missouri's successful use of similar aggregates in asphalt surface mixtures and base aggregate applications.

Aggregate Soundness Tests

Recent studies have focused on identifying alternatives to current tests to better predict aggregate durability. One of the main tests being considered is the Micro-Deval. A 2006 study in Wisconsin examined ways to "improve the effectiveness and cost-efficiency of the Wisconsin Department of Transportation's (WisDOT) aggregate durability testing protocol." *(Weyers, et.al., 2005)* Seventy-four aggregate types were tested according to nine test methods, including:

- AASHTO TP 58-00 – Micro-Deval
- CSA A23.2-24A – Unconfined Freezing and Thawing
- ASTM C131-01 – L.A. Abrasion Test
- ASTM C127-01 and ASTM C 128-01 (modified) – Vacuum Saturated Absorption

The results of the study suggested that the best performing test methods were the Micro-Deval test, the vacuum saturation test, and the L.A. Abrasion test. It was believed that the Micro-Deval test better represented the degradation experienced during mixing and handling. It also recommended using the Unconfined Freezing and Thawing test over the Sodium Sulfate Soundness test because it yielded stronger precision and better represented field performance.

Another independent study examined 23 aggregate sources subjected to the Micro-Deval, L.A. Abrasion, sodium sulfate soundness, and magnesium sulfate soundness tests. Upon completion, the Micro-Deval test results were compared with the other tests. Even though a good correlation was found between the sodium and magnesium sulfate soundness tests, no significant correlation was found between the Micro-Deval test and the two sulfate soundness tests or the L.A. Abrasion test. *(Rangaraju, 2008)*

A recent study by the Montana Department of Transportation, which currently uses the sodium sulfate soundness test for aggregate durability, investigated the use of the Micro-Deval test as an alternative. The department conducted the Micro-Deval, L.A. Abrasion, and sodium sulfate tests on a group of aggregates with various durability levels. Upon completion, the experimenters recommended the Micro-Deval test as a replacement for the sodium sulfate test, provided that another durability test was performed to support the results. *(Cuelho, et.al., 2007)*

A 2002 study in Iceland used aggregate from 20 different gravel pits and quarries to measure degradation of aggregates. Tests were conducted in three different categories: fragmentation, weathering, and abrasion. The weathering tests included three different freeze-thaw tests and the magnesium sulfate soundness test. Abrasion tests included the L.A. Abrasion test and the Micro-Deval. The test results were put through an analysis calculation using Varimax rotation, which is a statistical tool used in factor analysis. When all the data was plotted on the circle by test type, high correlation was found within each testing group. The study concluded then that it was not necessarily important which test was conducted for each group. It is worth noting that the majority of the aggregates used in this test were basaltic. The researchers acknowledged that results may be different for sedimentary, plutonic, or metamorphic rock. *(Bjarnason, et.al., 2002)*

Aggregate Soundness and Performance

Additional research has been performed in an attempt to relate aggregate properties to pavement performance. One experiment performed in Hawaii, where aggregates are significantly different than other U.S. locations, considered the relationship of the L.A. Abrasion test to long term pavement performance. Aggregates from each of twelve quarries in Hawaii were tested according to AASHTO T 96, but the results were said to correlate poorly with field performance, and a recommendation was made to replace the L.A. Abrasion method. The aggregate durability index (AASHTO T 210) was considered as a substitute. Next, the relationship between the magnesium sulfate and sodium sulfate tests was investigated, and the results showed that for Hawaii's climate, the magnesium sulfate test provided the better relationship to pavement performance. *(Brandes and Robinson, 2006)*

In a study by the Texas Transportation Institute *(Martin, et.al., 2007)*, durability and soundness tests were conducted on 16 U.S. aggregate sources to determine which method best related to field performance. Durability and soundness tests included the sodium and magnesium sulfate soundness tests (AASHTO T 104), Aggregate Freezing and Thawing (AASHTO T 103), Aggregate Durability Index (AASHTO T 210), Canadian Freeze-Thaw, L.A. Abrasion (AASHTO T 96), and the Micro-Deval test (AASHTO T 327). Upon conclusion of the testing regimen, it was determined that the L.A. Abrasion Test and Sodium Sulfate Soundness tests did not relate well to pavement performance at all. The tests that best related to field performance were the Micro-Deval and magnesium sulfate soundness tests. *(Yildirim, et.al., 2006)*

Wheel Tracking and Soundness Tests

The Texas Department of Transportation (TxDOT) has been successfully using the Hamburg Wheel-Tracking Device (HWTD) as part of their mixture design specification for several years. Each mixture must be gyratory compacted and then subjected to a LWT to determine whether it meets the specification before it can be approved for use. TxDOT has maintained a database of all test results. In 2006, a study was conducted by TxDOT and FHWA to determine if the HWTD could validate aggregate durability tests. Testing variables included mixture type (B, C, and D), aggregate type (gravel, igneous, and limestone-dolomite), binder type (PG 64-22, PG 70-22, and PG 76-22), testing temperature (40°C and 50°C), and mixture additives (none, lime, and liquid antistrip). The response variables used in the analysis included number of wheel passes and average deformation. The Micro-Deval and magnesium sulfate soundness tests were compared to the HWTD data, separated by passing and failing the HWTD. The results of the analysis indicated that the Micro-Deval and magnesium sulfate soundness tests did not relate well to the Hamburg test results. The researchers cited two probable reasons for this. First, more dominant variables (including binder type and temperature) influenced the HWTD results and masked the effects of aggregate durability. Second, additional aggregate characteristics, such as angularity, shape and texture, may have had a significant influence. It was suggested that additional research include mixtures where binder type and test temperature were held constant so that aggregate characteristics could be varied. *(Wu, et.al., 1998)*

Wheel Tracking and Pavement Performance

In another Texas study, the HWTD was compared to pavement performance data in order to determine whether or not a significant relationship was present. The research included monitoring the construction of test sections and monitoring performance over a 5-year period, then comparing the field data to laboratory test data gathered from HWTD testing. Nine aggregate types were selected and used in three different 12.5mm Superpave mix designs with PG 76-22 binder. The test section included portions of the eastbound and westbound lanes of Interstate Highway 20. HWTD testing was performed on laboratory mixtures and field cores, and traffic data was also obtained. *(Yildirim and Stokoe, 2006)*

The results suggested that rutting in the field was significantly less than that in the HWTD, and no field sections exhibited stripping. As a result, no data was available to relate field stripping performance to laboratory stripping performance. One informative finding from the project, however, was that an average of 37 wheel passes represented one Equivalent Single Axle Load (ESAL). *(Yildirim and Stokoe, 2006)*

5. Objectives

The overall objective of this project was to evaluate various methods for testing the soundness and durability of aggregates used in the construction of flexible and rigid pavements. Specific objectives were to:

- *Conduct a comprehensive examination of current literature regarding methods for measuring aggregate durability and soundness*. There are numerous test methods available for measuring the durability and soundness of aggregates. Thus the existing available literature was reviewed in order to gather information regarding each method, including advantages and disadvantages, and typical specification limits associated with each. Special attention was given with regard to the variability of each test method, including accuracy, precision, repeatability, and reproducibility. Relationships of aggregate soundness and pavement performance were sought, as were existing specifications.
- *Investigate various methods for the measurement of aggregate durability and soundness*. Laboratory testing was performed in order to quantify aggregate durability for selected aggregates. Several test methods were chosen and evaluated with respect to variability, cost, testing time, subjectivity, and procedural difficulties. Comparisons of alternative methods to those currently specified by AHTD were made.
- *Determine the relationships of each measure of aggregate durability and soundness to pavement performance*. Concrete and asphalt samples were prepared using mixtures containing the selected aggregate sources. Next, the samples were conditioned to simulate the environmental conditions affecting an in-place pavement. Then, performance testing was performed on the mixtures so that significant relationships between aggregate soundness and pavement performance could be identified.
- *Recommend a test method for inclusion in the current construction specification*. Of the methods investigated, the one(s) providing the strongest relationship to pavement performance should be considered for use. However, test method reliability is also an important factor. Thus, given all advantages and disadvantages of each soundness test method, the most beneficial method was chosen for incorporation into AHTD specifications.

6. Research Approach and Analysis

In this study, a variety of aggregate sources were characterized, tested for soundness properties by several methods, then used in HMA and PCC pavement mixtures and tested according to appropriate laboratory performance measures. In particular, carbonate aggregates from the northern sections of Arkansas were of interest because aggregate suppliers often contend that the current soundness tests do not provide accurate indications of the performance capabilities of limestone and dolomite aggregates. In order to contrast the carbonate aggregates, a syenite (i.e., non-carbonate) aggregate source was also included.

The primary purpose of the testing program was to evaluate the soundness characteristics of various carbonate aggregates, with additional focus on the variability of the results of the soundness tests. The primary performance characteristic in question was the ability of each aggregate source to perform under environmental and weathering conditions, specifically the freeze-thaw resistance of each aggregate.

Aggregate Selection

In this project, eight different aggregate sources were selected for testing. These sources represented four general locations in the state of Arkansas (as shown in Figure 9), and three different aggregate mineralogies. The aggregate types included limestone, dolomite, and syenite.

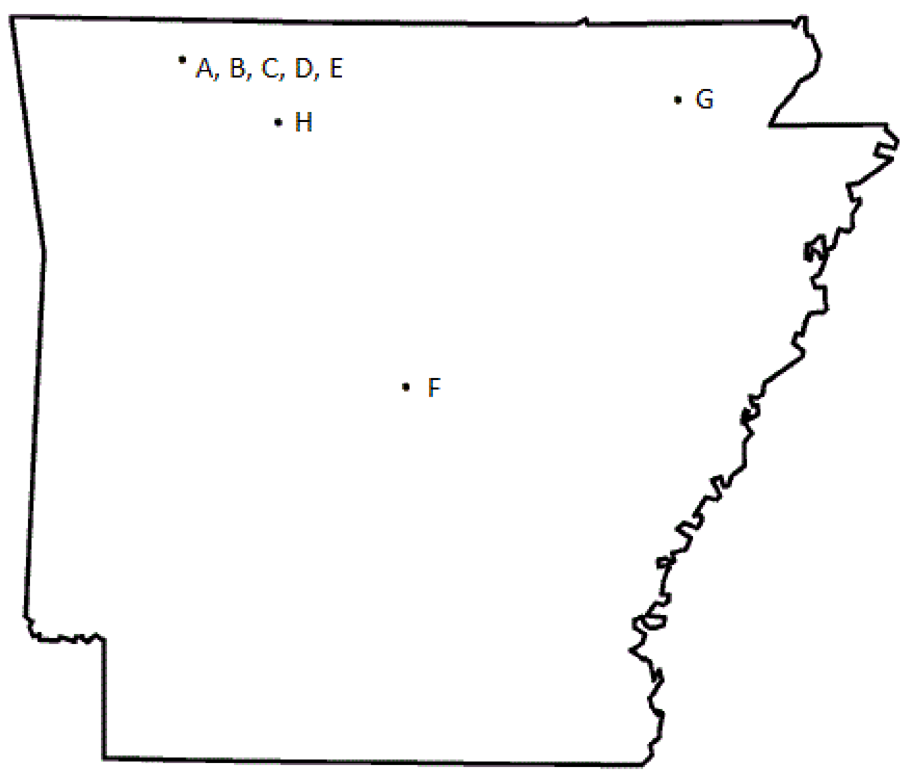

Figure 9. Locations of Aggregate Sources

Limestone is a sedimentary carbonate rock that contains a minimum of 50 percent calcium carbonate by weight and has a hardness of 4 on a Mohs Hardness Scale. It has many different uses, and various types exist. Limestone is generally used in the construction industry because it is strong and dense with few pore spaces, causing it to resist abrasion and freeze-thaw damage.

Dolomite rock is a sedimentary carbonate rock usually produced from a limestone rock which has been altered into dolomite by altering the mineral calcite. Dolomite is the second most abundant of the carbonate minerals and is used as a building material and a source of magnesium for the chemical industry. It has a hardness rating of 4.5 to 5 on a Mohs Hardness Scale, and a typical specific gravity of approximately 2.85.

Unlike limestone and dolomite, syenite is a rare, coarse-grained dense igneous rock composed mainly of feldspars, mica, hornblende, and pyroxene. Syenite is also similar in appearance and composition to granite; however, it has little or no quartz. Syenite has a hardness of about 6 on a Mohs Hardness Scale, and has very low absorption capacity. Of the three mineralogies used in this project, it possesses the highest density and lowest absorption capacity.

The eight aggregate sources used in this project are identified as A, B, C, D, E, F, G, and H. Aggregate sources A, B, C, D, and E are dolomite aggregates from the Berryville, Arkansas area, and are from varying ledges of two pits (termed 'old' and 'new') within the quarry. Aggregate F is a syenite aggregate source from central Arkansas, Aggregate G is a dolomite from the north eastern portion of Arkansas, and Aggregate H is a limestone from north central Arkansas. Specific descriptions follow, and aggregate source rankings are shown in Table 2. These general rankings are based on performance histories and anecdotal accounts of experiences associated with each material, and were treated as 'known' levels of performance for comparison purposes.

- Aggregate A came from a ledge of the new pit, and has been used in both HMA and PCC pavements. At one time, this aggregate source was approved for use by AHTD, but is no longer. This aggregate was considered to be of marginal overall quality.
- Aggregate B came from a ledge in the old pit, and has also been used for both HMA and PCC pavements. Similar to Aggregate A, this material was once approved for use by AHTD, but is no longer. This aggregate was characterized as marginal to poor in quality.
- Aggregate C came from a separate ledge in the old pit, and is somewhat similar to Aggregate B. This material is not currently approved by AHTD, and is characterized as marginal to poor in quality.
- Aggregate D came from a ledge in the new pit, and is currently approved for use by AHTD for use in both HMA and PCC pavement materials. Although it is approved for use, a limited number of individual samples exhibited losses that exceeded specification limits. The overall quality ranking of this aggregate is considered marginal.
- Aggregate E came from an upper ledge in the old pit, and is not considered to be of good quality. This aggregate meets the AHTD gradation requirements for Class 8 base rock, but has

never been approved for use in HMA or PCC paving materials. Aggregate from this bench sometimes contains a seam of dirt or clay, and is most often used for county work. The overall quality ranking of this aggregate is poor.

- Aggregate F was a syenite material from central Arkansas, and is the only aggregate source in the project that is a non-carbonate material. The syenite material has a history of high quality and is approved by AHTD for use in HMA and PCC paving materials. Because it is the only non-carbonate material tested, this aggregate source was treated as the control source for the study, and was categorized as a good quality aggregate source.
- Aggregate G was a dolomite from the Vulcan Quarry at Black Rock in the northeastern portion of Arkansas. This aggregate is currently approved for use by AHTD and is considered to be of good quality.
- Aggregate H was a limestone from the APAC (formerly McClinton Anchor) Quarry at Valley Springs in northern Arkansas. This material is approved by AHTD for use in both HMA and PCC paving materials, and is ranked in this study as having good quality.

Table 2. Aggregate Ranking

Aggregate	Quality Ranking
F	Good
G	
H	
D	
A	Marginal
C	
B	
E	Poor

Aggregate Characterization

The first task in the laboratory study involved a characterization for each of the eight aggregate sources. Approximately one ton of aggregate was obtained from each source, allowing for all project testing on each aggregate to be performed on materials acquired from a single sampling event. This strategy, combined with appropriate representative sampling and reducing techniques, minimized the potential for variability within each aggregate source. The test methods shown in Table 3, currently specified by AHTD, were performed for each aggregate source in triplicate. Standard test procedures were used for each method.

Table 3. Aggregate Characterization Tests

Method	Description
AASHTO T 2	Sampling of Aggregates
AASHTO T 11	Percent Finer than the No. 200 Sieve in Mineral Aggregate by Washing
AASHTO T 27	Sieve Analysis of Aggregate
AASHTO T 84	Specific Gravity and Absorption of Fine Aggregate
AASHTO T85	Specific Gravity and Absorption of Coarse Aggregate
AHTD 302	Deleterious Materials
AHTD 303	Crushed Particles
AHTD 306	Total Insoluble Residue in Coarse Aggregate
ASTM D 4791	Flat and Elongated Particles
AASHTO T 21	Organic Impurities

Gradation

The first test performed was sieve analysis, including test methods AASHTO T 11 and AASHTO T 27. Gradation data is shown in Table 4, where each value represents the average of three replicate tests. In some cases, a finer gradation could be a sign of a weaker aggregate that is prone to breakdown, which could be evaluated by considering the percentage passing the #4 sieve. In addition, weaker aggregates could be more likely to produce additional fines during production, detected by an increase in the percentage passing the #200 sieve. However, the gradation testing does not include any environmental conditioning, and did not consistently reflect known aggregate quality. Although in some cases the gradation could indicate the potential of an aggregate to break down, the actual gradation is much more significantly affected by the crushing operation and desired gradation for the source. Thus, no practically significant relationships were noted. Rankings based on the #4 and #200 sieves are compared to the known rankings in Table 5.

Table 4. Gradation results for each aggregate source, average of three results

	Average Percent Passing (%)							
Sieve	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**
1-1/2"	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1"	100.0	100.0	100.0	100.0	100.0	75.8	92.3	80.1
¾"	77.2	88.9	89.2	83.5	94.0	57.1	83.9	65.1
½"	40.9	65.6	63.8	54.3	76.2	38.5	71.8	50.8
3/8"	24.9	47.8	44.5	29.1	62.4	32.0	65.0	43.9
#4	4.7	11.8	8.8	0.9	38.3	24.1	46.5	32.2
#8	0.8	1.4	1.4	0.3	25.0	18.4	30.4	24.2
#16	0.7	1.0	1.2	0.2	16.7	15.5	21.2	18.9
#30	0.6	0.9	1.2	0.2	12.7	13.2	16.4	15.0
#50	0.6	0.8	1.1	0.2	10.2	11.0	13.2	11.2
#100	0.6	0.8	1.1	0.2	8.9	9.1	10.3	8.2
#200	0.6	0.7	1.0	0.2	7.6	7.4	7.7	6.2

Table 5. Aggregate source rankings based on gradation

Known Rank		**%Passing #4 Rank**	**%Passing #200 Rank**
F		D	D
G		A	A
H		C	B
D		B	C
A		F	H
C		H	F
B		E	E
E		G	G

Specific Gravity and Absorption

Next, specific gravity tests were performed according to AASHTO T 84 and AASHTO T 85. The coarse aggregate specific gravity was tested for all eight aggregate sources, while the fine aggregate specific gravity was only determined for those aggregates having a significant portion (i.e., more than 10 percent) of fine aggregate. This was true for Aggregates B, E, F, G and H. For the aggregates tested by both methods, the volumetric proportioning calculation described in AASHTO T 84 was used to arrive at the final values. A summary of the specific gravity and absorption values is given in Table 6, which includes apparent specific gravity, bulk specific gravity, bulk specific gravity with SSD basis, and absorption capacity.

Table 6. Specific gravity and absorption for each aggregate source, average of three results

Specific Gravity and Absorption Values

	A	B	C	D	E	F	G	H
Apparent Sp. Gr.	2.798	2.811	2.809	2.797	2.792	2.630	2.806	2.698
Bulk Sp. Gr.	2.621	2.569	2.641	2.660	2.652	2.577	2.718	2.641
Bulk (ssd) Sp. Gr.	2.684	2.655	2.701	2.708	2.702	2.598	2.749	2.662
Absorption (%)	2.4	3.4	2.3	1.9	1.9	0.8	1.2	0.8

At first consideration, it would seem that aggregate sources with low specific gravities (i.e., densities) and high absorption capacities could be more susceptible to environmental effects because of their increased ability to take on water that could freeze and expand within the aggregate pores. The specific gravity values were ranked and compared to the known rankings to see if a relationship was evident. These rankings are shown in Table 7.

Table 7. Aggregate source rankings based on specific gravity and absorption

Known Rank		Apparent Sp. Gr. Rank	Bulk Sp. Gr. Rank	Bulk (ssd) Sp. Gr. Rank	Absorption (%) Rank
F		B	G	G	F
G		C	D	D	H
H		G	E	E	G
D		A	H	C	D
A		D	C	A	E
C		E	A	H	C
B		H	F	B	A
E		F	B	F	B

From these rankings, it appears that of these measures, the absorption capacity of the aggregate source is best able to predict known quality, while specific gravity does not provide a reasonable prediction of aggregate performance. It was expected that density would not properly characterize the performance, but that absorption could provide some insight. Aggregate sources that are able to take on more water could retain that water during freezing weather, at which time the expansion forces within the aggregate pores could be great enough to damage the structural integrity of the aggregate particles. Aggregate pore size is also a factor that could affect the behavior of an aggregate during freezing and thawing, as larger pores would allow a quicker release of absorbed water than smaller pores; however, this characteristic is more difficult to measure. Based on the absorption values shown and known performance, a maximum absorption value of 2.0 percent could indicate that more intensive soundness testing is warranted.

Aggregate Shape

Aggregate shape is an important feature in the performance of an aggregate in a paving mixture. Shape is most critical for asphalt pavements because the aggregate structure forms the skeleton of the mixture and is the primary provider of mixture strength. Crushed, cubical, and angular aggregates tend to increase the level of aggregate interlock and generate additional mixture strength. Flat or elongated particles can interfere with consolidation and result in materials that are difficult to place during construction, and may be more prone to degradation during production. In this study, each aggregate source was tested according to AHTD 304, 'Crushed Particles in Aggregate', and ASTM D 4791 'Flat Particles, Elongated Particles, or Flat and Elongated Particles in Coarse Aggregate'. Average results are given in Table 8. Because all of the aggregates chosen were composed of crushed quarry rock, all sources were believed to be 100 percent crushed, which was confirmed in the testing. With regard to flat and elongated, small percentages were determined to be flat, very minor percentages were elongated, and only two sources had extremely small percentages of both flat and elongated. It is noted that all testing was performed using a 1:2 ratio, which is much more conservative and detected a greater percentage of flat and/or elongated particles than if the 1:5 ratio had been used. The 1:5 ratio is typically used in Arkansas. Although aggregate shape is not intuitively related to its resistance to freeze/thaw damage, a more cubical aggregate has less surface area and fewer potential surface voids to absorb water than a flat and/or elongated particle.

Table 8. Aggregate Shape Data, average values

Aggregate Shape Values

	A	B	C	D	E	F	G	H
Crushed Particles, %	100	100	100	100	100	100	100	100
Flat Particles, %	3.27	2.19	1.82	2.87	1.15	8.14	3.62	2.66
Elongated Particles, %	0.40	0.40	0.57	0.29	0.09	1.80	2.16	1.00
Flat & Elongated Particles, %	0.00	0.00	0.00	0.00	0.00	0.17	0.17	0.00

Aggregate Impurities

Organic impurities can cause a detrimental effect on the strength of mortar in concrete, and is often used in making a preliminary decision to accept or reject fine aggregate used in concrete paving materials. Likewise, deleterious materials such as slate, shale, clay lumps, and friable particles can affect their performance of an aggregate source. Each of the eight aggregate sources were tested in triplicate according to AASHTO T 21, 'Organic Impurities in Fine Aggregates for Concrete', and AHTD Method 302 'Deleterious Matter in Aggregate'. Only one sample of the three samples tested from aggregate source B was found to contain deleterious material, measured at 1.03 percent, which is well under the 5 percent limit according to AHTD. None of the samples tested exhibited any concern with regard to organic impurities.

Aggregate Soundness

The next portion of the project involved a thorough investigation of various soundness tests. Each aggregate source was tested according to traditional soundness tests, including sodium sulfate soundness by AASHTO T 104, magnesium sulfate soundness by AASHTO T 104, aggregate freeze-thaw by AASHTO T 103, and the Micro-Deval Abrasion test by AASHTO T 327. Additional non-standard test methods were also investigated and performed, including a modified freeze-thaw test, vacuum saturation, and the SG-9 aggregate specific gravity and absorption test.

<u>Sodium Sulfate Soundness</u>

The sodium sulfate test was performed on triplicate samples of each of the eight aggregate sources according to AASHTO T 104, using a five cycle freeze-thaw sequence. The results are given in Table 9, and examples of aggregate deterioration resulting from this test method are shown in Figure 10. The average results were relatively low, with only Aggregate B exceeding the AHTD specification limit of 12 percent. However, three individual test results exceeded the limit. Variability data is also shown in Table 9, including standard deviation and coefficient of variation (COV). In general, it is desirable for COV values to be no more than approximately 15 percent, however the average COV for the sodium sulfate soundness test was 57 percent – a very high value, indicating poor repeatability for this test method. While COV can be a key indicator of variability, it is often more beneficial to identify the exact sources of variability. Thus, an additional statistical analysis was performed for the complete dataset to determine the overall proportion of variability within the experiment. The total variability in the experiment was separated to quantify the percentage of variability that could be attributed to differences between the aggregate sources and the pure error of the experiment, which represents the variability of the test method. Of the total experimental variability, 36.8 percent was attributed to the differences among aggregate sources, while 63.2 percent was inherent in the test method. It is noteworthy that the aggregate sources chosen for the project were intended to represent a wide range of soundness characteristics, and the method itself generated almost twice the variability of the aggregate sources. In other words, the unintentional variability was approximately twice that of the intentional experimental variation.

Table 9. Sodium Sulfate Soundness test results and variability data

Sodium Sulfate Soundness Test Results

	A	B	C	D	E	F	G	H
$NaSO_4$ Loss (%)	4.94	10.55	6.68	3.91	4.22	0.52	2.38	0.90
	12.95	22.82	14.52	4.35	5.65	0.74	2.17	1.65
	2.80	7.67	5.73	10.36	11.14	1.37	7.05	1.84
Average Loss, %	6.90	13.68	8.98	6.21	7.00	0.88	3.87	1.46
Standard Deviation	5.35	8.05	4.82	3.60	3.65	0.44	2.76	0.50
COV, %	77.6	58.8	53.7	58.06	52.16	50.32	71.35	34.0
Variability due to aggregate source, %						36.8		
Variability due to test method, %						63.2		

Figure 10. Aggregates After Testing by the Sulfate Soundness Method (AASHTO T 104)

Next, the sodium sulfate soundness results were used to rank aggregate quality. A comparison of those rankings with the known rankings is given in Table 10. It appears that the sodium sulfate test, despite its obvious issues with variability, was able to reasonably rank the aggregates.

Table 10. Aggregate source rankings based on sodium sulfate soundness

Known Rank		$NaSO_4$ Rank
F		F
G		H
H		G
D		D
A		A
C		E
B		C
E		B

<u>Magnesium Sulfate Soundness</u>

The magnesium sulfate test was also performed on triplicate samples of each of the eight aggregate sources according to AASHTO T 104, using the five cycle freeze-thaw sequence. The results are given in Table 11. The average results are significantly higher than those for the sodium sulfate method, with four of the aggregates having an average result that exceeded a typical specification limit of 18 percent. Thirteen of the 24 individual test results exceeded the 18 percent limit. Variability data is also shown in Table 11, including standard deviation and coefficient of variation (COV). The average COV for the magnesium sulfate soundness test was approximately 20 percent – slightly higher than desired, yet much better than for the sodium sulfate counterpart. In terms of assigning sources of variability, the magnesium sulfate test was able to limit the pure error of the test method to 5.5 percent, while the other 94.5 percent could be explained by actual differences between the aggregate sources. This indicates that the test method was much more capable of producing repeatable results and detecting actual differences in aggregate soundness characteristics. In this case, the intentional variability was much greater than the unplanned experimental variation.

Table 11. Magnesium Sulfate Soundness test results and variability data

Magnesium Sulfate Soundness Test Results

	A	B	C	D	E	F	G	H
$MgSO_4$ Loss (%)	15.25	37.50	28.66	28.56	33.46	4.21	12.07	6.04
	20.82	46.42	24.61	23.06	30.77	1.79	6.04	4.83
	9.02	47.59	29.85	24.44	34.62	3.26	9.75	4.59
Average Loss, %	15.03	43.84	27.71	25.35	32.95	3.09	9.29	5.15
Standard Deviation	5.90	5.52	2.75	2.86	1.98	1.22	3.04	0.78
COV, %	39.3	12.6	9.9	11.3	6.0	39.5	32.8	15.1
Variability due to aggregate source, %						94.5		
Variability due to test method, %						5.5		

Next, the magnesium sulfate soundness results were used to rank aggregate quality. A comparison of those rankings with the known rankings is given in Table 12. It appears that this test was able to adequately rank the aggregates.

Table 12. Aggregate source rankings based on magnesium sulfate soundness

Known Rank		$MgSO_4$ Rank
F		F
G		H
H		G
D		A
A		D
C		C
B		E
E		B

Micro-Deval Abrasion

The Micro-Deval Abrasion test was performed on triplicate samples of the eight aggregate sources according to AASHTO T 327. The results are given in Table 13. Typical Micro-Deval requirements allow a maximum percent loss ranging from 15 to 25 percent. *(Martin, et.al., 2007)* Assuming a cut-off value of 20 percent, all eight aggregate sources would be considered acceptable based on average test results, although one individual test result slightly exceeded this limit. If a cut-off value of 15 percent was used, which is more typical of aggregates used in surface paving mixtures, four of the eight aggregates would have been deemed unacceptable based on average test results, with 13 of 24 individual test results exceeding the limit. Variability data is shown in Table 13, including standard deviation and coefficient of variation (COV). The average COV for the Micro-Deval test was approximately 8 percent, which is an acceptable value. In terms of assigning sources of variability, the Micro-Deval was able to limit the pure error of the test method to 4.4 percent, while the other 95.6 percent could be explained by actual differences between the aggregate sources. This indicates that the test method was much more capable of producing repeatable results and detecting actual differences in aggregate soundness characteristics. Again, the intentional variability was much greater than the unplanned experimental variation.

Table 13. Micro-Deval Abrasion test results and variability data

Micro-Deval Test Results

	A	B	C	D	E	F	G	H
Micro-Deval Loss (%)	16.96	17.84	14.52	12.38	16.86	6.23	11.14	19.90
	16.03	19.47	14.39	12.71	14.54	5.04	9.00	20.31
	15.57	19.51	13.69	15.28	14.45	4.60	10.62	19.66
Average Loss, %	16.19	18.94	14.20	13.46	15.28	5.29	10.25	19.96
Standard Deviation	0.71	0.95	0.45	1.59	1.37	0.84	1.12	0.33
COV, %	4.4	5.0	3.1	11.8	8.9	15.9	10.9	1.6
Variability due to aggregate source, %					95.6			
Variability due to test method, %					4.4			

Next, the Micro-Deval data was used to rank aggregate quality. A comparison of those rankings with the known rankings is given in Table 14. Despite the reduced variability of this test method, the rankings were not very consistent with the known performance levels. Two of the good performers were accurately detected, but marginal and poor performers were inconsistent. Particularly, Aggregate H was known to be one of the better performers, but was considered the worst performer according to the Micro-Deval. It is important to consider that the Micro-Deval is an abrasion test and does not include any environmental cycling or freeze/thaw conditioning aside from the fact that the test is performed with water. Thus, this test method may be a better indicator of the aggregate's performance with respect to polishing and degradation from external forces than the environmental effects of freezing and thawing (i.e., soundness).

Table 14. Aggregate source rankings based on Micro-Deval

Known Rank		Micro-Deval Rank
F		F
G		G
H		D
D		C
A		E
C		A
B		B
E		H

Aggregate Freeze-Thaw

Aggregate freeze-thaw testing was performed on triplicate samples of the eight aggregate sources according to AASHTO T 103, using total immersion as outlined in Procedure A, a 3 percent NaCl and water solution, and 50 cycles. All testing for the AASHTO T 103 method was performed using a computer-controlled freeze-thaw chamber to automatically produce the temperature cycle such that the temperature of the specimen was reduced to -23°C (-9°F) and held for two hours, then raised to 21°C (70°F) and held for 30 minutes. The results are given in Table 15, and photographs of aggregates tested by this method are shown in Figure 11. This method is only used in a few states, and typical requirements limit the maximum percent loss to 18 percent. Based on average test results, two aggregates (B and C) would have failed this requirement, and one (Aggregate E) would be considered marginal. Five of the 24 individual test results exceeded the limit. Variability data is also shown in Table 15, including standard deviation and coefficient of variation (COV). The average COV for the aggregate freeze-thaw test was approximately 36 percent, which is somewhat excessive. In terms of variability sources, the pure error of the test method accounted for over half of the total variability, indicating that the test method was approximately as variable as the test method itself. There was essentially as much intentional variability as there was unintentional variability.

Table 15. Aggregate Freeze-Thaw test results and variability data

Aggregate Freeze-Thaw Test Results

	A	B	C	D	E	F	G	H
Freeze-Thaw Loss (%)	6.82	10.12	13.36	10.89	16.58	5.65	13.52	2.10
	16.07	49.69	36.73	14.19	16.77	5.65	14.97	1.34
	9.08	27.46	23.29	12.14	18.05	1.86	8.01	1.37
Average Loss, %	10.66	29.09	24.46	12.41	17.13	4.39	12.17	1.60
Standard Deviation	4.82	19.84	11.73	1.67	0.80	2.19	3.67	0.43
COV, %	45.3	68.2	48.0	13.4	4.7	49.9	30.2	26.8
Variability due to aggregate source, %						46.7		
Variability due to test method, %						53.3		

Figure 11. Aggregates After Testing by the Aggregate Freeze-Thaw Method (AASHTO T 103)

Next, the freeze-thaw data was used to rank aggregate quality. A comparison of those rankings with the known rankings is given in Table 16. The relatively high level of variability associated with this test method is consistent with its inability to rank the aggregates, as the rankings were not very consistent with the known performance levels. The aggregate freeze-thaw method ranked most of the aggregates fairly well by very general categories, but ranked Aggregate H (limestone) better than Aggregate F (the non-carbonate aggregate).

Table 16. Aggregate source rankings based on Aggregate Freeze-Thaw (AASHTO T 103)

Known Rank		T 103 Rank
F		H
G		F
H		A
D		G
A		D
C		E
B		C
E		B

<u>Aggregate Freeze-Thaw by Deep Freeze Method</u>

The general idea behind the aggregate freeze-thaw test is to simulate freezing and thawing in the field by placing samples in a controlled environment with accelerated temperature cycles. This process requires expensive, specialized equipment, and is not readily available to many laboratories. As a surrogate, an alternative method for simulated freeze-thaw testing of aggregates was developed. This method, called the Aggregate Freeze-Thaw by Deep Freeze Method, utilized a standard, residential-grade chest freezer to produce the low temperature for the freeze cycle (i.e., approximately 4°F), and a controlled temperature environment at room temperature (i.e., approximately 70°F) for the thawing cycle. The aggregate specimens were prepared as outlined in AASHTO T 103, and tested in total immersion using a 0.5 percent isopropyl alcohol and water solution. Each aggregate sample was placed in solution in a plastic container with a lid, and soaked for 24±4 hours at room temperature. Next, the sample containers were placed in the deep freeze for 24±2 hours, then moved to a location at room temperature for 24±2 hours. This 48 hour sequence of freeze and thaw constituted a complete freeze-thaw cycle. Aggregate samples were subjected to five freeze-thaw cycles, and then sieved as described in AASHTO T 103 to determine percent loss. If, at any time, the test had to be interrupted, the sample was covered and maintained in a thawed state until testing was resumed. Although fewer cycles were used by this method than the traditional T 103, the cycles were longer, allowing for a more realistic speed of freezing and thawing. A detailed procedure is provided in Appendix A. The results of the tests are shown in Table 17, and examples of aggregate distress resulting from this test method are shown in Figure 12.

Table 17. Aggregate Freeze-Thaw by Deep Freeze test results and variability data

Aggregate Freeze-Thaw by Deep Freeze (DF) Test Results

	A	B	C	D	E	F	G	H
Freeze-Thaw (DF) Loss (%)	12.64	29.68	27.84	6.95	17.18	1.75	12.05	1.10
	9.11	23.36	11.82	5.44	27.12	1.35	14.28	1.28
	11.20	14.75	13.31	7.53	12.77	1.33	14.02	0.96
Average Loss, %	10.98	22.60	17.66	6.64	19.02	1.48	13.45	1.11
Standard Deviation	1.77	7.49	8.85	1.08	7.35	0.24	1.22	1.11
COV, %	16.2	33.2	50.1	16.2	38.6	16.0	9.1	14.4
Variability due to aggregate source, %						70.0		
Variability due to test method, %						30.0		

Figure 12. Aggregates After Testing by the Deep Freeze Method

Because this method is not a standard and is not currently used by agencies, there are no specification limits associated with its use. However, it is intended to provide information similar to that of AASHTO T

103, so an appropriate limit for loss would be 18 percent. Based on average test results, two aggregates (B and E) would have failed this requirement, and one (Aggregate C) would be considered marginal. Four of the 24 individual test results exceeded the limit. Variability data is also shown in Table 17, including standard deviation and coefficient of variation (COV). The average COV for the aggregate freeze-thaw test was approximately 24 percent, which is somewhat excessive, but better than the AASHTO T 103 COV. In terms of variability sources, the pure error of the test method accounted for 30 percent, while the actual differences in aggregate type comprised the other 70 percent. This indicates that there was approximately twice as much intentional variability as there was unintentional variability. Thus, in terms of variability, the deep freeze method was more repeatable than the standard AASHTO T 103 method.

Next, the freeze-thaw data was used to rank aggregate quality. A comparison of those rankings with the known rankings is given in Table 18. The aggregate freeze-thaw by deep freeze method ranked most of the aggregates fairly well by very general categories, with most aggregates ranked within one position of the known rank. Aggregate H was ranked somewhat higher than its known rank, and Aggregate G was ranked as mediocre rather than good. The Deep Freeze rankings were not vastly different from the rankings by AASHTO T 103, with both methods ranking Aggregate H (limestone) as the highest in quality.

Table 18. Aggregate source rankings based on Aggregate Freeze-Thaw by Deep Freeze

Known Rank		Deep Freeze Rank
F		H
G		F
H		D
D		A
A		G
C		C
B		E
E		B

The primary advantage of this test method is that if it is proven to be accurate and repeatable, it could provide valuable soundness information for aggregate sources using equipment and materials that are more readily available to most laboratories. In terms of testing time, the total length of testing for AASHTO T 103 was approximately 3 weeks, while the average testing time for the Deep Freeze method was just two weeks. One disadvantage of the Deep Freeze method, however, is that a laboratory technician must be available at approximately the same time each day to transfer the samples from the freezer to room temperature (or vice-versa); the AASHTO T 103 method used the freeze-thaw chamber which is automatically controlled, and no technicians were needed to interact with the samples during the entire sequence of freeze-thaw cycles.

The variability of the deep freeze method was significantly lower than that of the AASHTO T 103 method, indicating that the deep freeze method could adequately serve as a surrogate for AASHTO T 103. However, neither test method was significantly better than the magnesium sulfate method.

Vacuum Saturation

Another testing variation that was developed was the vacuum saturation test of coarse aggregates. The goal of this test is to determine whether placing a coarse aggregate sample under vacuum would affect its measure of specific gravity and absorption capacity. AASHTO T 85 requires a 15 to 19 hour soaking period in order to determine absorption capacity, and it is generally assumed that the pore spaces of most aggregates will be completely filled after that length of time. However, aggregates that do not become essentially saturated during that time are likely to have higher actual absorption capacities, and may be more susceptible to taking on water in the field, thereby exacerbating freeze-thaw distress.

For the vacuum saturation test, each aggregate sample was prepared according to AASHTO T 85, except that after being oven dried, cooled, and weighed, the aggregate was placed in water and 25 to 30 mm Hg of vacuum was applied to the specimen for 5 minutes. After slowly removing the vacuum, the sample was removed from the water and brought to the SSD condition as described in AASHTO T 85. The SSD weight was recorded, and then a submerged weight of the specimen was determined. Then the specimen was again submerged and placed under vacuum for 10 minutes (giving the aggregate a total cumulative time of 15 minutes under vacuum), and then SSD and submerged weights recorded again. Another 15 minutes of vacuum was then applied to the sample, producing a cumulative time of 30 minutes under vacuum, and SSD and submerged weights were recorded again. Finally, the specimen was soaked in water for an additional 20 to 24 hours, then measured again for SSD and submerged weights, and dried to a constant mass to determine a final dry weight. Specific gravity and absorption values were calculated for each time interval according to AASHTO T 85. A data summary showing average values for each measured response is given in Table 19.

Table 19. Vacuum Saturation average test results

Vacuum Saturation Test Results (average of 3 values)

	A	B	C	D	E	F	G	H
Bulk Sp. Gr., 5min	2.613	2.587	2.642	2.636	2.612	2.615	2.754	2.649
Bulk Sp. Gr., 15min	2.615	2.591	2.651	2.638	2.614	2.621	2.759	2.654
Bulk Sp. Gr., 30min	2.616	2.590	2.654	2.638	2.616	2.622	2.761	2.662
Bulk Sp. Gr., 24hr	2.624	2.621	2.640	2.612	2.613	2.623	2.763	2.654
Absorption %, 5min	2.3	2.7	1.9	1.8	2.2	0.1	0.7	0.7
Absorption %, 15min	2.7	3.2	2.2	2.1	2.7	0.4	1.2	0.9
Absorption %, 30min	2.7	3.2	2.1	2.1	2.8	0.3	1.2	0.8
Absorption %, 24hr	2.7	2.8	2.4	2.7	2.9	0.4	1.1	0.8

In most cases, the specific gravity and absorption increased with increasing time under vacuum. The increase in absorption appeared to be most pronounced for Aggregates B, C, D, and E, which was reasonable because those aggregates also exhibited the greatest T 85 absorption capacities. Aggregate B, exhibiting the greatest overall absorption capacity, actually experienced a decrease in absorption between the 30 minute vacuum and the 24 hour soak. This could be explained if the pore spaces in the aggregate were large enough that a vacuum were strong enough to hold water in some of the pores, while a simple soak was not, and allowed some absorbed water to escape from those pores. In other cases, as with Aggregate D, the absorption capacity sharply increased after the 24-hour soak, meaning that additional time allowed more water to enter the smaller and inner-most pores. Although some variability is likely attributed to variations in identifying the SSD condition consistently, this is evidence that pore size could play a critical role in aggregate soundness behavior.

In terms of rankings, absorption capacity was much more capable than bulk specific gravity of correctly ranking the aggregates. A comparison of those rankings with the known rankings is given in Table 20. In terms of absorption, the best performer (i.e., lowest percent absorption) was Aggregate F, the non-carbonate aggregate. All measures of absorption correctly identified the good aggregates (F, G, and H) in the top three positions, and fairly consistently identified Aggregates B and E as the poor performers. In general, absorption capacity appears capable of ranking aggregates, such that absorption capacities exceeding 2 percent could be an indicator that further soundness testing is necessary.

Table 20. Aggregate source rankings based on Vacuum Saturation Test Results

Known Rank		BSG, 5min	BSG, 15min	BSG, 30min	BSG, 24hr	ABS% 5min	ABS% 15min	ABS% 30min	ABS% 24hr
F		G	G	G	G	F	F	F	F
G		H	H	H	H	G	H	H	H
H		C	C	C	C	H	G	G	G
D		D	D	D	A	D	D	C	C
A		F	F	F	F	C	C	D	A
C		A	A	E	B	E	A	A	D
B		E	E	A	E	A	E	E	B
E		B	B	B	D	B	B	B	E

Next, a statistical comparison was made between the AASHTO T 85 and vacuum saturation measures of absorption. In a single-factor analysis of variance (treating aggregate source as a blocking factor), the method of measurement (i.e., time of soak or vacuum) produced a significant difference. Further analysis was performed according to Duncan's method for means testing, which revealed that absorption capacity based on the 5 minute vacuum produced the smallest value, followed by that for the T 85 method. The largest value was associated with the 24 hour soak after the 30 minute vacuum. Table 21 displays the statistical results, indicating statistical significance for a 95 percent level of significance. For each absorption measure, the mean response is given. Means that do not display statistically significant differences are denoted by a solid underline. For instance, the responses for T 85, ABS30m and ABS15m are connected by a solid underline, indicating that there is not a significant difference for measures obtained from AASHTO T 85, and the 15 minute and 30 minute vacuum saturation periods. The 5 minute vacuum time was significantly less than the other measures, suggesting that the aggregate pores were not essentially filled after the 5-minute vacuum period. Thus, this measure was eliminated from further analysis. Because similarities were seen among the other measures, the T85 and ABS24h measures of absorption were selected for further analysis of variability composition.

Table 21. Statistical Comparison of Various Measures of Absorption (%)

	ANOVA / Means Test Results Response = Absorption, %				
significant	ABS5m	T85	ABS30m	ABS15m	ABS24h
p < 0.0001	1.56	1.81	1.90	1.92	1.98

Individual test results, averages, and variability data are shown in Table 22 for absorption values measured according to AASHTO T 85 and vacuum saturation with a 24 hour soak. In general, COV percentages indicate that both methods are decently consistent, with an average COV of 6.6 percent for AASHTO T 85, and 9 percent for the vacuum saturation method. In terms of variability composition, AASHTO T 85 exhibited only 1.8 percent unintentional variability, and over 98 percent attributed to differences in the aggregate source. The vacuum saturation procedure was almost as adept, with 4 percent of the error associated with the method, and 96 percent due to actual differences in aggregate source. Thus, absorption capacity is certainly a parameter that is capable of differentiating between various levels of quality, and is relatively capable of separating good and fair/poor quality aggregates.

Table 22. Absorption test results and variability data

Aggregate Absorption Capacity Test Results

	A	B	C	D	E	F	G	H
Absorption by T 85 (%)	2.32	3.30	2.46	1.79	1.86	0.67	1.03	0.82
	2.56	3.34	2.21	1.88	1.92	0.83	1.41	0.77
	2.33	3.42	2.14	1.90	1.90	0.76	1.01	0.80
Average, %	2.40	3.35	2.27	1.86	1.89	0.75	1.15	0.80
Standard Deviation	0.14	0.06	0.17	0.06	0.03	0.08	0.23	0.03
COV, %	5.6	1.8	7.4	3.2	1.6	10.6	19.6	3.2
Variability due to aggregate source, %						98.2		
Variability due to test method, %						1.8		
Absorption by 24hr (%)	2.73	3.34	2.45	2.67	3.24	0.27	1.16	0.77
	2.72	2.56	2.48	2.60	2.68	0.43	1.09	0.94
	2.72	2.52	2.40	2.72	2.69	0.45	1.17	0.81
Average, %	2.72	2.81	2.44	2.66	2.87	0.38	1.14	0.84
Standard Deviation	0.01	0.46	0.04	0.06	0.32	0.10	0.04	0.09
COV, %	0.2	16.5	1.7	2.3	11.2	25.7	3.8	10.6
Variability due to aggregate source, %						96.0		
Variability due to test method, %						4.0		

<u>SG-9</u>

The SG-9 Aggregate Specific Gravity and Absorption Device is marketed by Gilson as an alternative to AASHTO T-85, providing fast and accurate measures of specific gravity and absorption in a simple manner. This device, shown in Figure 13, includes an acrylic sample chamber, a calibrated measuring vessel, a displacement sensor, and a laptop computer with preloaded software. For coarse aggregates, the sample chamber, containing approximately 1000 grams of aggregate is lowered into the measuring vessel. For fine aggregates, the sample is poured into a measuring tube while a stirring device distributes the sample evenly. A series of measurements are made during a 5-minute period, and calculations are performed to determine bulk specific gravity and absorption without the need for visually identifying the saturated surface dry condition. Software prompts guide the user through the process, and display the results immediately after completion of the test.

Figure 13. The SG-9 Aggregate Specific Gravity & Absorption Device

This device was used to test Aggregates A and B in order to determine whether this method would warrant further study. A summary of results is given in Table 23. Bulk specific gravity values measured by the SG-9 were somewhat higher than those measured by AASHTO T 85. Average SG-9 bulk specific gravity measures for Aggregates A and B were 2.743 and 2.749, respectively, and were 2.621 and 2.569 by AASHTO T 85. Absorption values measured by the SG-9 were generally lower than those by T 85, with average SG-9 absorption capacities of 1.6 and 2.7 for Aggregates A and B, respectively. Values by T 85 were 2.40 and 3.35. Regarding variability, the coefficient of variation was very low for specific gravity, however most of the overall variability exhibited (97.2 percent) was due to the test method and not aggregate source. For absorption, the coefficient of variation was considerably higher, but a much larger portion of the total variability could be attributed to actual differences in aggregate source.

The specific gravity measurements obtained by the SG-9 method appeared to be quite repeatable, especially for Aggregate B. This method warrants further investigation for making such determinations. Absorption values by the SG-9 were more variable than those by T 85, making the SG-9 a somewhat less desirable method than T 85 for quantifying absorption capacity. Because absorption capacity appeared to relate more closely to aggregate soundness than the measures of specific gravity, this method was not included in further study for this project. However, its use is recommended for additional study for measuring specific gravity.

Table 23. SG-9 test results and variability data

SG-9 Aggregate Specific Gravity and Absorption Capacity Test Results

	A	B
SG-9 Bulk Specific Gravity	2.679	2.750
	2.783	2.749
	2.767	2.749
Average	2.743	2.749
Standard Deviation	0.049	0.007
COV, %	1.8	0.3
Variability due to aggregate source, %		2.8
Variability due to test method, %		97.2
SG-9 Absorption (%)	1.11	2.64
	1.78	2.59
	1.98	2.72
Average, %	1.62	2.65
Standard Deviation	0.402	0.216
COV, %	24.8	8.1
Variability due to aggregate source, %		78.0
Variability due to test method, %		22.0

Variability

The first portion of the investigation of soundness test methods focused on testing variability, the composition of testing variability, and the ability of each to rank aggregate performance correctly based on levels of known performance. The test methods with the least variability, as measured by coefficient of variation, were absorption capacity by AASHTO T 85 (6.6 percent), the Micro-Deval method (7.7 percent), and the 24-hour absorption by vacuum saturation (9.0 percent). Of the more traditional soundness methods, all had COV values exceeding 20 percent. Of those methods, the COV values were: magnesium sulfate soundness (20.8 percent), aggregate freeze-thaw by deep freeze (24.2 percent), aggregate freeze-thaw by AASHTO T 103 (35.8 percent), and sodium sulfate soundness (57.0 percent).

In terms of the composition of that variability, those displaying the least amount of pure error (i.e., unintentional variability) were T 85 absorption (1.8 percent), absorption by vacuum saturation (4.0 percent), Micro-Deval (4.4 percent), and magnesium sulfate soundness (5.5 percent). Thus the most advantageous methods based on variability were the Micro-Deval, magnesium sulfate soundness, and absorption measures.

Accuracy

Accuracy was assessed based on the comparisons of rank by a particular test method to the known rank. In terms of these rankings, the most capable methods of correctly ranking aggregate performance were sodium sulfate soundness, magnesium sulfate soundness, aggregate freeze-thaw by deep freeze, T 85 absorption, and absorption by vacuum saturation. Although the sodium sulfate soundness was one of the better predictors, this method was also one of the most variable methods, making its true accuracy questionable.

Discrimination

Next the analysis considered the ability of each method to discern actual differences between aggregate sources. Since the aggregates were chosen with the express purpose of demonstrating varying levels of quality, these differences should be detected in order for the test methods to be considered effective measures of aggregate quality. An analysis of variance (ANOVA) was performed for each test method to determine statistically whether each method could clearly distinguish among the aggregates of varying quality. The results of the ANOVA, based on a 95 percent level of significance, are given in Table 24. All methods were able to detect some level of significance between the aggregate sources, although this ability was only marginal for the sodium sulfate soundness test. This raises considerable concern regarding the AHTD specification because there is a large difference in actual aggregate known performance, but this was difficult to consistently detect by the sodium sulfate method. In other words, the variability associated with the test method masked the true differences between aggregate sources.

Table 24. Discrimination of Test Methods (ANOVA results)

Test Method	p-value	Significant?
Sodium Sulfate Soundness	0.0447	marginal
Magnesium Sulfate Soundness	<0.0001	Yes
Micro-Deval	<0.0001	Yes
Aggregate Freeze-Thaw (AASHTO T 103)	0.0155	Yes
Aggregate Freeze-Thaw (Deep Freeze)	0.0003	Yes
Aggregate Absorption (AASHTO T 85)	<0.0001	Yes
Aggregate Absorption (24-hr vacuum saturation)	<0.0001	Yes

Mixture Performance

In the next portion of the study, each of the 8 aggregate sources was used to generate asphalt and concrete paving mixtures, and applicable performance tests were performed on each mixture type.

HMA Designs

Because most of the aggregate sources had a nominal maximum aggregate size (NMAS) of 25.0mm, HMA mixtures were designed to meet the gradation criteria for 25.0mm NMAS Superpave mixtures. For aggregate sources containing larger aggregate particles, the aggregate source was pre-screened over the

1-inch sieve to remove the coarser fraction. For most of the aggregate sources, a single aggregate source did not satisfy the gradation requirements, so an additional (finer) aggregate source was added to achieve a reasonable gradation. In most cases, the additional source consisted of sandstone. Carbonate aggregates are often combined with sandstone aggregates in order to meet the siliceous materials requirement for surface courses. Although the designs prepared were binder course mixtures and not surface mixtures, it was believed that the addition of a sandstone material would best represent aggregate combinations most commonly used mixture designs. All HMA mixtures were designed using a PG 70-22 polymer-modified binder, and compacted to 100 design gyrations, as nearly as possible to the guidelines of the AHTD Standard Specification. In reality, HMA mixtures are produced using several aggregate components. However, the aggregates used in the job mix formula were limited as much as possible to the primary aggregate type in question so that performance characteristics could be attributed to the primary aggregate source, and not confounded by aggregate mixtures. Due to this, some deviations were allowed during the mix design process. No anti-strip products were used for the mix designs. A summary of each mixture is given in Table 25.

Table 25. HMA Mix Design Summary

	A	B	C	D	E	F	G	H
NMAS	25.0	25.0	25.0	25.0	25.0	25.0	25.0	25.0
PG	70-22	70-22	70-22	70-22	70-22	70-22	70-22	70-22
N_{des} (gyr)	100	100	100	100	100	100	100	100
	Job Mix Formula (%)							
Primary Aggregate	66	70	70	66	100	70	95	60
Sandstone Screenings	34	30	30	34				
Syenite Screenings						30		
Natural Sand							5	
¾” Sandstone								20
½” Limestone								20
	Blend Gradation (%Passing)							
1-1/2”	100	100	100	100	100	100	100	100
1”	100	100	100	100	100	93	96	96
¾”	85	92	92	89	94	82	89	89
½”	61	76	75	70	76	74	72	69
3/8”	50	64	61	53	62	68	59	53
#4	33	36	33	31	38	57	34	31
#8	19	19	18	19	25	43	24	20
#16	14	13	13	13	17	28	17	13
#30	12	11	11	11	13	15	14	11
#50	10	9	9	10	10	11	9	8
#100	6	3	6	6	9	9	6	6
#200	4.4	2.6	4.2	4.1	5.0	6.9	3.8	4.5
	Volumetric Properties							
Binder Content (%)	4.4	4.5	4.6	4.4	4.4	5.8	4.6	3.4
Air Voids (%)	4.5	4.5	4.9	4.3	4.4	4.5	4.5	4.4
VMA (%)	13.4	12.5	14.5	14.8	13.1	16.1	13.6	13.2
VFA (%)	66.4	64.0	66.2	70.9	66.4	72.0	66.9	66.7
Gsb	2.62	2.57	2.64	2.66	2.65	2.58	2.72	2.64
Gse	2.662	2.640	2.672	2.650	2.703	2.616	2.782	2.638
Gmm	2.487	2.465	2.487	2.477	2.541	2.405	2.580	2.481

<u>HMA Performance</u>
The performance of each of the HMA mixtures was determined using three tests:

- The Evaluator of Rutting and Stripping in Asphalt (ERSA)
- AASHTO T 283, Resistance of Compacted Hot Mix Asphalt to Moisture Induced Damage
- TxDOT Method, TEX-245-F, Test procedure for Cantabro Loss

These tests were chosen because a mixture that is susceptible to moisture damage would almost certainly experience a greater acceleration of damage if aggregate particles were broken due to freeze-thaw cycles. The broken, exposed faces would allow moisture to enter the aggregate/asphalt interface, and further exacerbate layer deterioration. Also, mixtures that are prone to moisture damage are more likely to allow moisture to penetrate the mix layer, enhancing the damage done by the freeze-thaw cycles. The Cantabro test is essentially a measure of mixture toughness, but was believed to also have potential for identifying mixture performance.

The ERSA testing was performed in the submerged state at a temperature of 50°C, with a wheel load of 132 pounds. Each test continued until 20,000 cycles of the loaded wheel had been applied, or until a maximum rut depth (approximately 20 mm) was reached. The response variables produced from each ERSA test included rut depth at 10,000 cycles (Rut10k, mm), rut depth at 20,000 cycles (Rut20k), rutting slope (RSlope), stripping slope (SSlope), and stripping inflection point SIP. Two subsets of ERSA specimens from each aggregate source were tested: one set was tested after preparation in the gyratory compactor, and the second set was conditioned using one freeze-thaw cycle prior to ERSA testing. A summary of ERSA test results are given in Table 26.

Table 26. ERSA Test Results

	Unconditioned Specimens					Conditioned with 1 F/T Cycle				
	Rut10k (mm)	Rut20k (mm)	RSlope (cyc/mm)	SSlope (cyc/mm)	SIP (cyc)	Rut10k (mm)	Rut20k (mm)	RSlope (cyc/mm)	SSlope (cyc/mm)	SIP (cyc)
A	7.62	15.66	1280	835	13243	5.44	8.87	4320	2735	9712
A	7.36	14.04	3218	1508	9204	9.68	11.89	1147	469	3636
B	18.22	18.20	484	375	1123	15.89	16.50	794	470	1255
B	21.18	21.24	459	177	1633	16.74	16.78	625	385	2966
C	14.32	17.66	979	499	5860	14	18.50	1145	504	4307
C	14.18	19.72	1410	639	5606	13.50	18.50	1410	639	4456
D	8.63	13.19	2278	2278	DNS*	9.72	15.07	1735	1735	DNS*
D	12.09	21.44	964	693	9637	10.84	17.59	1311	709	7325
E	11.12	18.28	1318	777	8160	10.61	15.91	1930	876	4624
E	15.66	21.26	963	874	8955	6.62	10.04	2658	1944	5827
F	8.36	11.50	2339	2339	DNS*	9.81	15.58	1668	1668	DNS*
F	18.45	20.00	753	324	4425	8.93	16.88	1936	821	15935
G	4.78	6.28	10267	3013	17300	7.08	10.91	5187	1227	17306
G	8.99	12.48	2925	2925	DNS*	10.50	16.65	1515	704	11080
H	8.06	16.90	1576	728	13516	6.52	14.10	1879	1066	14165
H	7.79	17.60	1581	622	11197	10.86	20.18	1531	682	9472

**DNS = Did not strip*

Next, the data from each response variable was used to rank the aggregate sources. These results are shown in Table 27. In general, the stripping inflection point was most accurate at ranking the aggregate sources, particularly for specimens conditioned with a freeze-thaw cycle prior to testing. However, none of the responses tended to clearly mimic the known rankings of the aggregate sources. While a clear trend would have been desirable, it is reasonable that a number of other factors could affect the rutting and stripping performance of the mixtures, leading to additional variability in the rankings.

Table 27. Aggregate source rankings based on ERSA Test Results

Known Rank	Unconditioned Specimens					Conditioned with 1 F/T Cycle				
	Rut10k (mm)	Rut20k (mm)	RSlope (cyc/mm)	SSlope (cyc/mm)	SIP (cyc)	Rut10k (mm)	Rut20k (mm)	RSlope (cyc/mm)	SSlope (cyc/mm)	SIP (cyc)
F	G	G	G	G	G	A	A	G	A	F
G	A	A	A	D	D	E	E	A	E	D
H	H	F	D	F	F	H	G	E	F	G
D	D	H	H	A	H	G	F	F	D	H
A	E	D	F	E	A	F	D	H	G	A
C	F	C	C	H	E	D	B	D	H	E
B	C	B	E	C	C	C	H	C	C	C
E	B	E	B	B	B	B	C	B	B	B

AASHTO T 283 was also used to predict the long-term stripping susceptibility of the HMA mixes. In this test, a set of gyratory-compacted specimens with 6-inch diameters was prepared for each of the 8 mixtures. Then, the conditioned specimens were vacuum saturated to a level of 70 to 80 percent, subjected to a 24-hour freeze cycle, then placed in a hot water bath prior to breaking in indirect tension as described in AASHTO T 283. A second set of samples was prepared for each of the 8 mixtures, from which a 4-inch diameter specimen was cored and used for testing. Coring was performed in order to expose aggregate faces and create a harsher simulation of field conditions, creating the potential to more accurately assess the impact of aggregate quality on HMA laboratory performance. The results of the moisture damage testing are given in Table 28. It is noted that AHTD specifications require AHTD method 455, which is similar to the AASHTO T 283 method, but utilizes a Marshall-style fixture and does not require a freeze-thaw cycle.

Table 28. Moisture Damage (AASHTO T 283) Test Results

	6"-diameter compacted specimens			4"-diameter cored specimens		
	Dry Tensile Strength (psi)	**Wet Tensile Strength (psi)**	**Tensile Strength Ratio, (%)**	**Dry Tensile Strength (psi)**	**Wet Tensile Strength (psi)**	**Tensile Strength Ratio, (%)**
A	223.1	184.6	80.3	361.2	267.6	75.7
A	210.0	168.0		374.1	224.9	
A	217.3	169.2		303.4	283.4	
B	233.1	164.8	65.3	307.8	225.5	63.7
B	257.2	151.9		343.1	199.2	
B	238.0	156.8		290.3	173.0	
C	296.7	206.4	68.7	411.3	289.4	69.0
C	311.5	224.2		394.3	283.6	
C	311.5	198.1		387.8	252.4	
D	325.0	128.4	67.3	360.9	273.8	73.0
D	296.5	204.4		355.1	272.6	
D	257.1	237.9		398.7	264.7	
E	326.0	176.2	55.3	312.4	274.2	68.0
E	317.2	178.7		364.2	235.4	
E	279.9	156.4		389.4	199.4	
F	253.0	202.4	89.0	268.6	232.1	84.0
F	213.1	194.5		287.7	244.4	
F	210.5	202.2		287.9	233.6	
G	273.6	181.7	72.7	291.0	228.7	70.0
G	264.6	198.2		321.2	211.9	
G	213.1	164.6		307.7	199.8	
H	294.6	203.4	80.3	301.5	201.3	69.0
H	300.4	202.6		348.0	229.1	
H	184.8	194.7		347.1	256.1	

Most specifications require a minimum tensile strength ratio of 80 percent. Based on these results, few of the aggregate sources produced mixtures that met this criterion. Only Aggregate F (the known best performer) was clearly successful, while Aggregates A and H were marginally successful. In terms of rankings, dry tensile strength, wet tensile strength, and tensile strength ratio for the 6-inch specimens and 4-inch cores were used to rank the aggregate sources, as shown in Table 29. Dry measures of tensile strength were not indicative of actual performance, as the best known performer was ranked poorly for both sample sets. However, since the dry measures did not include any sample conditioning, these values were not expected to reflect the known soundness performance. The wet tensile strength measures were slightly more accurate at matching the known rankings, but not considered to be accurate predictors of performance. The tensile strength ratios were most capable of predicting soundness performance, with deviations noted for Aggregates A and D.

Table 29. Aggregate source rankings based on AASHTO T 283 Test Results

Known Rank	6"-diameter compacted specimens			4"-diameter cored specimens		
	Dry Tensile Strength (psi)	Wet Tensile Strength (psi)	Tensile Strength Ratio, (%)	Dry Tensile Strength (psi)	Wet Tensile Strength (psi)	Tensile Strength Ratio, (%)
F	E	C	F	C	C	F
G	C	H	A	D	D	A
H	D	F	H	E	A	D
D	H	D	G	A	F	G
A	G	G	C	H	E	H
C	B	A	D	B	H	C
B	F	E	B	G	G	E
E	A	B	E	F	B	B

The next test method used to ascertain the relative performance of the HMA mixtures was the Texas Cantabro Loss test. This method is typically used for open-graded friction courses (OGFC), or porous friction courses (PFC), with percent loss usually being limited to 20 percent. *(Martin, et.al., 2007)* Since the samples tested in this project were not OGFCs, the 20 percent limit would be considered excessive. This test did, however, provide a measure of the integrity of the bond between the aggregates and binder. While this test does not include any conditioning cycles and is not necessarily based on environmental factors, it does evaluate the ability of the aggregates and binder to maintain a tight bond while the specimen is under distress. The results and rankings of the Cantabro Loss test are given in Table 30. The Cantabro test was able to adequately identify Aggregate F as the best performer, and Aggregate E as a poor performer; however, many of the other aggregates were not ranked correctly.

Table 30. Cantabro Loss Test Results and Rankings

Aggregate	% Loss
A	12.36
B	7.89
C	9.85
D	9.73
E	11.78
F	4.99
G	9.51
H	7.37

Known Rank	Cantabro Rank
F	F
G	H
H	B
D	G
A	D
C	C
B	E
E	A

In addition to rankings, the performance test methods were also evaluated to determine their abilities to discern between varying aggregate quality. An ANOVA was performed for each of the methods to determine which response variables could detect significant differences between aggregate type. The results of these analyses are shown in Table 31, and are based on a 95 percent level of significance.

Table 31. ANOVA Summary of Discrimination for HMA Performance Response Variables

Response Variable	p-value	Significant?
ERSA – Unconditioned		
Rut Depth at 10,000 cycles, mm	0.0409	Marginal
Rut Depth at 20,000 cycles, mm	0.2112	No
Rutting Slope, cyc/mm	0.2001	No
Stripping Slope, cyc/mm	0.0656	Marginal
Stripping Inflection Point, cycles	0.4355	No
ERSA Conditioned		
Rut Depth at 10,000 cycles, mm	0.0281	**Yes**
Rut Depth at 20,000 cycles, mm	0.2026	No
Rutting Slope, cyc/mm	0.5369	No
Stripping Slope, cyc/mm	0.7274	No
Stripping Inflection Point, cycles	0.2522	No
AASHTO T 283 – 6-inch Specimens		
Dry Tensile Strength, psi	0.0120	**Yes**
Wet Tensile Strength, psi	0.1387	No
Tensile Strength Ratio (TSR)	0.1149	No
AASHTO T 283 – 4-inch Cores		
Dry Tensile Strength, psi	0.0012	**Yes**
Wet Tensile Strength, psi	0.0116	**Yes**
Tensile Strength Ratio (TSR)	0.4110	No
Cantabro Loss, %	0.2339	No

Based on rankings, the ERSA Stripping Inflection Point was the most successful parameter at judging aggregate soundness. However, this parameter was not statistically capable of discerning between varying levels of aggregate quality when these aggregates were used in HMA mixtures. This was true for both the conditioned and unconditioned sample sets. In fact, the only ERSA parameter that was able to significantly distinguish aggregate quality was rut depth at 10,000 cycles for the conditioned sample set. The only moisture damage response for the 6-inch specimens that could discriminate between aggregate type was the dry tensile strength, which was least similar to the known rankings. For the 4-inch cores, wet and dry tensile strengths were best able to discern aggregate quality. The Cantabro test was not able to distinguish between the aggregate sources.

An additional analysis was performed on the ERSA data to assess the significance of sample conditioning. A single factor completely randomized block design (block = aggregate source) was used to determine whether sample conditioning was a significant factor on the responses of rut depth at 10,000 cycles, rut depth at 20,000 cycles, rutting slope, stripping slope, and stripping inflection point. Determinations were based on a 95 percent level of significance, and are shown in Table 32. In no case was the conditioning process significant. This means that the ERSA data was not significantly affected by the addition of a single freeze-thaw conditioning cycle, and that the datasets can be combined in future analyses. One explanation for the lack of effect could be that during mixing, aggregates are coated with asphalt binder, which is intended to be impermeable. If the binders perform properly, no moisture will enter the pores of the aggregate, and thus, no significant damage will be caused. So, it is reasonable that a significant effect of sample conditioning would only be evident after the integrity of the binder coatings has been damaged.

Table 32. ANOVA Results – Significance of Sample Conditioning for ERSA Testing

Response Variable	p-value	Significant?
Rut Depth at 10,000 cycles, mm	0.1872	No
Rut Depth at 20,000 cycles, mm	0.2565	No
Rutting Slope, cyc/mm	0.8217	No
Stripping Slope, cyc/mm	0.6389	No
Stripping Inflection Point, cycles	0.5255	No

The same type of ANOVA was then performed for the moisture damage tests to determine the significance of coring each specimen to a 4-inch diameter before testing. The results are shown in Table 33. Tensile strength ratio was not significantly affected; however, the dry and wet tensile strengths were significantly affected by the process of coring a 6-inch specimen to produce a 4-inch diameter. Thus, the datasets could not be combined in further analyses.

Table 33. ANOVA Results – Significance of Sample Conditioning for Moisture Damage Testing

Response Variable	p-value	Significant?
Tensile Strength Ratio	0.7973	No
Dry Tensile Strength, psi	<0.0001	**Yes**
Wet Tensile Strength, psi	<0.0001	**Yes**

Overall, no HMA mixture performance parameter was capable of adequately distinguishing the varying levels of quality while also accurately ranking aggregate soundness performance. Again, there are likely a number of other mixtures properties that have confounded the performance test results, masking the relationships between aggregate quality and HMA mixture performance.

PCC Designs

The eight aggregate sources were also used to generate Portland Cement Concrete (PCC) mixture designs. For each mixture, aggregate source was used as the coarse aggregate component, and a natural sand (having a specific gravity of 2.600 and an absorption capacity of 0.3 percent) was used for the fine aggregate component. No other aggregate products were substituted for mixture adjustment purposes, and all mixture designs were created as similarly as possible in order to better isolate the effects of coarse aggregate type. Type 1 cement with a specific gravity of 3.100 was used for each design. Mix design summaries for the PCC mixtures are shown in Table 34.

Table 34. PCC Mix Design Summary

	A	B	C	D	E	F	G	H
Design Air, %	6	6	6	6	6	6	6	6
Water/cement ratio	0.42	0.42	0.42	0.45	0.42	0.39	0.42	0.42
	Blend Gradation (%Passing)							
1-1/2”	100	100	100	100	100	100	100	100
1”	100	100	100	100	100	86	96	89
¾”	89	94	94	91	97	76	91	81
½”	68	81	80	76	87	65	85	73
3/8”	59	71	70	63	80	62	81	69
#4	47	50	49	47	65	56	70	61
#8	41	41	41	43	54	49	57	53
	Batch Weights per yd^3							
Cement, lbs	650	650	650	650	650	650	650	650
Coarse Aggregate, lbs	1700	1700	1700	1600	1700	1750	1700	1700
Sand, lbs	1407	1394	1422	1468	1412	1363	1450	1393
Water, lbs	273	273	273	293	273	254	273	273
Air, percent	2	2	2	2	2	2	2	2

Concrete Performance

The performance of each of the concrete mixtures was determined using three tests:

- ASTM C 666 – Resistance of Concrete to Rapid Freezing and Thawing
- ASTM C 215 – Fundamental Transverse, Longitudinal, and Torsional Resonant Frequencies of Concrete Specimens
- Concrete Strength Using Freeze-Thaw Cycles

These tests were chosen because they represent the performance of concrete mixtures when exposed to freeze-thaw cycles. The ASTM test methods (ASTM C 666 and ASTM C 215) are advanced methods that are fairly standard in the industry for determining the effects of freeze-thaw cycles. The third test was developed during the project as a simplified method for exposing the concrete cylinders to 5 and 10 freeze-thaw cycles and measuring the compressive strength. Although this was not an established standard method, it was a way to provide a relative comparison of concrete mixture strengths with and without freeze-thaw conditioning.

The rapid freezing and thawing test was performed according to test methods ASTM C 666 and C 215 on concrete beams using a freeze-thaw chamber in the laboratory and 300 conditioning cycles. The resonant frequency of each beam was determined on the beams prior to conditioning, and then periodically during the series of freeze-thaw cycles. The durability factor was determined by dividing the frequency after a given number of cycles by the frequency at 0 cycles. The durability ratios at 120, 200, and 300 cycles are shown in Table 35.

Table 35. Durability Ratio Test Results

	Durability Ratio @ 120 cycles, %	Durability Ratio @ 200 cycles, %	Durability Ratio @ 300 cycles, %
A	89.3	71.5	56.0
A	86.8	73.8	61.7
B	102.3	91.0	61.1
B	99.0	92.3	85.7
C	94.6	70.3	0.0
C	88	82.5	52.7
D	95.6	86.1	0.0
D	93.5	80.9	0.0
E	97.9	81.7	0.0
E	98.0	80.2	77.9
F	100.3	88.8	96.9
F	99.3	85.1	93.7
G	99.2	93.1	99.1
G	94.5	94.1	83.4
H	88.3	70.2	76.6
H	90.3	67.2	79.0

The data for each durability ratio measure was used to rank the aggregate sources. The results are shown in Table 36. The ratios in the earlier portions of the test sequence did not appear to indicate correct rankings for the aggregates. However, the durability ratio at 300 cycles was a fair indicator of aggregate soundness, particularly identifying the better performers.

Table 36. Durability Ratio Rankings

Known Rank		Durability Ratio @ 120 cycles, %	Durability Ratio @ 200 cycles, %	Durability Ratio @ 300 cycles, %
F		B	G	F
G		F	B	G
H		E	F	H
D		G	D	B
A		D	E	A
C		C	C	E
B		H	A	C
E		A	H	D

Next, a surrogate test method for measuring a concrete mixture's ability to withstand freeze-thaw distresses was used. In this method, concrete cylinders were prepared, then subjected to a series of 24-hour freeze and 24-hour thaw cycles using the deep freezer. One subset of cylinders was subjected to 5 freeze-thaw cycles, and another was subjected to 10 freeze-thaw cycles. After the conditioning cycles, each cylinder was tested for compressive strength. The resulting strengths were compared to the 14-day and 28-day compressive strengths determined for each of the 8 mixture designs. Test results are shown in Table 37. Interestingly, in almost every case, the compressive strength after 5 freeze-thaw cycles was greater than the 14-day strength, meaning that the conditioned specimens displayed greater compressive strengths than the unconditioned ones. This is contrary to intuitive thought, but could be explained by the fact that during the freeze-thaw cycles, the concrete continued to cure, generating additional strength gain. Thus, the 14-day strengths were not optimal for judging specimen quality. For most aggregates, however, strengths decreased between the 28-day strengths and strengths after conditioning with freeze-thaw cycles. It was suspected that the difference in the 28-day strengths and the strengths after 10 freeze-thaw cycles would be the most informative response variable, representing the loss of strength caused by the 10 freeze-thaw cycles.

Table 37. Compressive Strength Using Freeze-Thaw Test Results

	14-day Compressive Strength, psi	28-day Compressive Strength, psi	Compressive Strength w/ 5 F/T cycles, psi	Compressive Strength w/ 10 F/T cycles, psi	28-day Strength – 10 F/T cycles, psi
A	5048	6269	5681	5858	411
A	4870	5866	5997	6039	-173
B	3165	4536	3709	4070	466
B	3573	4312	3621	3791	521
C	5191	6452	5680	5708	744
C	4685	5591	5500	5924	-333
D	4815	5827	4885	5680	147
D	4896	5290	--	5517	-227
E	5115	6032	5016	5321	711
E	4796	5613	5359	5266	347
F	5350	6271	5474	6254	17
F	5486	6351	6143	6423	-72
G	4111	5082	5020	4960	122
G	--	5149	4720	4846	303
H	4083	5345	4941	5145	200
H	4251	4884	4578	5137	-253

Each strength response variable was used to rank the aggregate sources. Rankings are given in Table 38. Strengths, in general, were not able to correctly rank the aggregate sources, except that Aggregate F was identified as a good performer and Aggregate B was identified as a poor performer. The response variable best able to rank aggregate soundness quality was the difference in 28-day strength and strength after 10 freeze-thaw cycles.

Table 38. Strength Rankings Using Freeze-Thaw Test Results

Known Rank	14-day Compressive Strength, psi	28-day Compressive Strength, psi	Compressive Strength w/ 5 F/T cycles, psi	Compressive Strength w/ 10 F/T cycles, psi	28-day Strength – 10 F/T cycles, psi
F	F	F	A	F	D
G	A	A	F	A	F
H	E	C	C	C	H
D	C	E	E	D	A
A	D	D	D	E	C
C	H	G	G	H	G
B	G	H	H	G	B
E	B	B	B	B	E

Next, the performance test response variables were tested using ANOVA to determine whether each was capable of discriminating between varying levels of quality. The results are given in Table 39. For the concrete test methods, all were able to discriminate effectively with the exception of the difference in 28-day strength and strength after 10 freeze-thaw cycles. Unfortunately, this measure was determined to be one of the better indicators in terms of ranking. The durability ratio at 300 cycles was marginal.

Table 39. ANOVA Summary of Discrimination for PCC Performance Response Variables

Response Variable	p-value	Significant?
Durability Ratio		
Durability Ratio at 120 cycles, %	0.0044	**Yes**
Durability ratio at 200 cycles, %	0.0011	**Yes**
Durability ratio at 300 cycles, %	0.0438	marginal
Compressive Strength		
14-day compressive strength, psi	0.0003	**Yes**
28-day compressive strength, psi	0.0048	**Yes**
Compressive strength after 5 freeze-thaw cycles, psi	0.0009	**Yes**
Compressive strength after 10 freeze-thaw cycles, psi	<0.0001	**Yes**
Difference in 28-day and 10 freeze-thaw cycle, psi	0.6013	No

As with the HMA performance data, although no response variable was clearly identified as having the ability to both rank aggregate performance correctly and to significantly discriminate between aggregate sources, other mixture properties could have generated sources of variability that masked the effects of aggregate quality.

Relationships of Aggregate Soundness and Performance

In the next portion of the analysis, regression techniques were used to seek correlations between aggregate soundness properties and mixture performance. First, aggregate soundness measures were compared to known performance, with numerical values given to each known performance rank (i.e., 1 through 8). Then, aggregate soundness measures were compared to HMA and PCC mixture performance measures for the purpose of identifying significant correlations.

Multiple linear regression procedures were used to determine the soundness measures that were most capable of predicting the soundness performance of the aggregates. Stepwise, backward, and Best R^2 techniques were used, and the resulting relationships and associated R^2 values are shown in Table 40. The single most significant predictor variable was also identified for each model. For stepwise regression, variables entered the model based on a significance of 0.1500. Variables retained in backward regression procedures were retained in the model at a 0.1000 level of significance. The numbers of variables chosen for the Best R^2 models were determined based on a 3 percent increase threshold. For the first analysis, known rank was chosen as the dependent variable, while the following parameters were selected as predictor variables:

- Percent loss by the sodium sulfate soundness procedure, AASHTO T 104 (SSPL)
- Percent loss by the magnesium sulfate soundness procedure, AASHTO T 104 (MSPL)
- Percent loss by the Micro-Deval abrasion method, AASHTO T 327 (MDV)
- Percent loss by the aggregate freeze-thaw test, AASHTO T 103 (FT103)
- Percent loss of aggregates in freeze-thaw using the deep freeze (DFRZ)
- Absorption capacity of coarse aggregate, AASHTO T 85 (ABS)

This analysis provided determinations of the aggregate soundness properties that were most able to produce known aggregate rankings. Percent loss by the magnesium sulfate method alone was able to describe 75 percent of the variability in rank values, making it the single best predictor. The relationships became stronger as percent loss by the Micro-Deval and Deep Freeze methods were added to the model, however the additional variables did not significantly increase the R^2 values.

Table 40. Regression Summary Using Aggregate Soundness to Predict Aggregate Rank

Regression Method	Significant Factors	Model	R^2
Stepwise	MSPL, MDV, DFRZ	RANK = -0.16 + 0.095(MSPL) + 0.140(MDV) + 0.064(DFRZ)	0.83
Backward	MSPL, MDV, DFRZ	RANK = -0.16 + 0.095(MSPL) + 0.140(MDV) + 0.064(DFRZ)	0.83
Best R^2	MSPL, MDV	RANK = 0.110 + 0.124(MSPL) + 0.133(MDV)	0.80
Most significant single factor	MSPL	RANK = 1.593 + 0.143(MSPL)	0.75

<u>HMA Performance</u>

Next, the same regression procedures were used to predict performance rank of the aggregates by various measures of HMA mixture performance. Because the ERSA conditioned and unconditioned test results did not yield statistically significant differences, these datasets were combined in the regression analyses. Predictor variables in these analyses included:

- Rut Depth at 10,000 Cycles (RUT10K)
- Rut Depth at 20,000 Cycles (RUT20K)
- Rutting Slope (RSLOPE)
- Stripping Slope (SSLOPE)
- Stripping Inflection Point (SIP)
- Tensile Strength Ratio (TSR)

Table 41 provides the subset regression summary for the relationship of HMA performance parameters and aggregate rank.

Table 41. Regression Summary Using HMA Performance to Predict Aggregate Rank

Regression Method	Significant Factors	Model	R^2
Stepwise	SIP, TSR	RANK = 11.470 – 0.0001(SIP) – 7.879(TSR)	0.47
Backward	RSLOPE, SSLOPE, SIP, TSR	RANK = 11.083 – 0.0008(RSLOPE) + 0.002(SSLOPE) – 0.0002(SIP) – 7.337(TSR)	0.62
Best R^2	RSLOPE, SSLOPE, SIP, TSR	RANK = 11.083 – 0.0008(RSLOPE) + 0.002(SSLOPE) – 0.0002(SIP) – 7.337(TSR)	0.62
Most significant single factor	SIP	RANK = 5.921 – 0.0001(SIP)	0.34

While a fairly decent correlation was developed for aggregate soundness measures to predict aggregate rank, none of the asphalt mixture performance parameters were particularly successful at correlating with aggregate rank. The best individual indicators were stripping inflection point (R^2 = 0.34), rut depth at 10,000 cycles (R^2 = 0.18), tensile strength ratio (R^2 = 0.15), and stripping slope (R^2 = 0.12). None of these predictors, however, provided significant relationships with known aggregate rank. Thus, additional analyses were performed to relate aggregate soundness characteristics to HMA performance measures. Results for the more significant relationships are given in Tables 42 through 44. The relationships were generally weak, with the percent loss by magnesium sulfate soundness, Micro-Deval, sodium sulfate soundness, and deep freeze methods providing the greatest link to HMA performance.

Table 42. Regression Summary Using Aggregate Soundness to Predict Rut Depth at 10,000 Cycles

Regression Method	Significant Factors	Model	R^2
Stepwise	SSPL, DFRZ	RUT10K = 7.776 + 0.305(SSPL) + 0.109(DFRZ)	0.39
Backward	SSPL	RUT10K = 8.601 + 0.396(SSPL)	0.34
Best R^2	SSPL, MDV, DFRZ	RUT10K = 9.930 + 0.357(SSPL) – 0.182(MDV) + 0.118(DFRZ)	0.42
Most significant single factor	SSPL	RUT10K = 8.601 + 0.396(SSPL)	0.34

Table 43. Regression Summary Using Aggregate Soundness to Predict Stripping Inflection Point

Regression Method	Significant Factors	Model	R^2
Stepwise	MDV, DFRZ	SIP = 35272 -1177.3(MDV) -422.6(DFRZ)	0.37
Backward	MDV, DFRZ	SIP = 35272 -1177.3(MDV) -422.6(DFRZ)	0.37
Best R^2	SSPL, MDV, DFRZ	SIP = 33952 – 434.9(SSPL) – 996.7(MDV) – 308.8(DFRZ)	0.40
Most significant single factor	MDV	SIP = 33741 – 1446.8(MDV)	0.27

Table 44. Regression Summary Using Aggregate Soundness to Predict Tensile Strength Ratio

Regression Method	Significant Factors	Model	R^2
Stepwise	MSPL, ABS	TSR = 0.770 – 0.008(MSPL) + 0.062(ABS)	0.33
Backward	MSPL, ABS	TSR = 0.770 – 0.008(MSPL) + 0.062(ABS)	0.33
Best R^2	MSPL, ABS	TSR = 0.770 – 0.008(MSPL) + 0.062(ABS)	0.33
Most significant single factor	MSPL	TSR = 0.817 – 0.005(MSPL)	0.29

Although the relationships were not capable of producing predictive equations for relating aggregate soundness and HMA mixture performance, these relationships do provide a sense of the parameters that are most closely related. Percent loss by the magnesium sulfate method was determined to be the most advantageous measure of aggregate soundness, and was also the most significant variable relating to tensile strength ratio. The conditioning process of the magnesium sulfate method is intended to mimic the actual stresses within an aggregate's structure, which should also affect the stripping performance of an HMA mixture. Thus, a reasonable connection exists for these methods, and additional regression procedures were employed to determine whether a non-linear single regression technique would better quantify the relationship. The linear and logarithmic relationships are shown in Figure 14. As percent loss by the magnesium sulfate method increased, the tensile strength ratio decreased, showing that decreased aggregate performance generates diminished HMA mixture performance. Although the logarithmic relationship increases the R^2 value, this increase was minimal, and did not significantly improve the relationship.

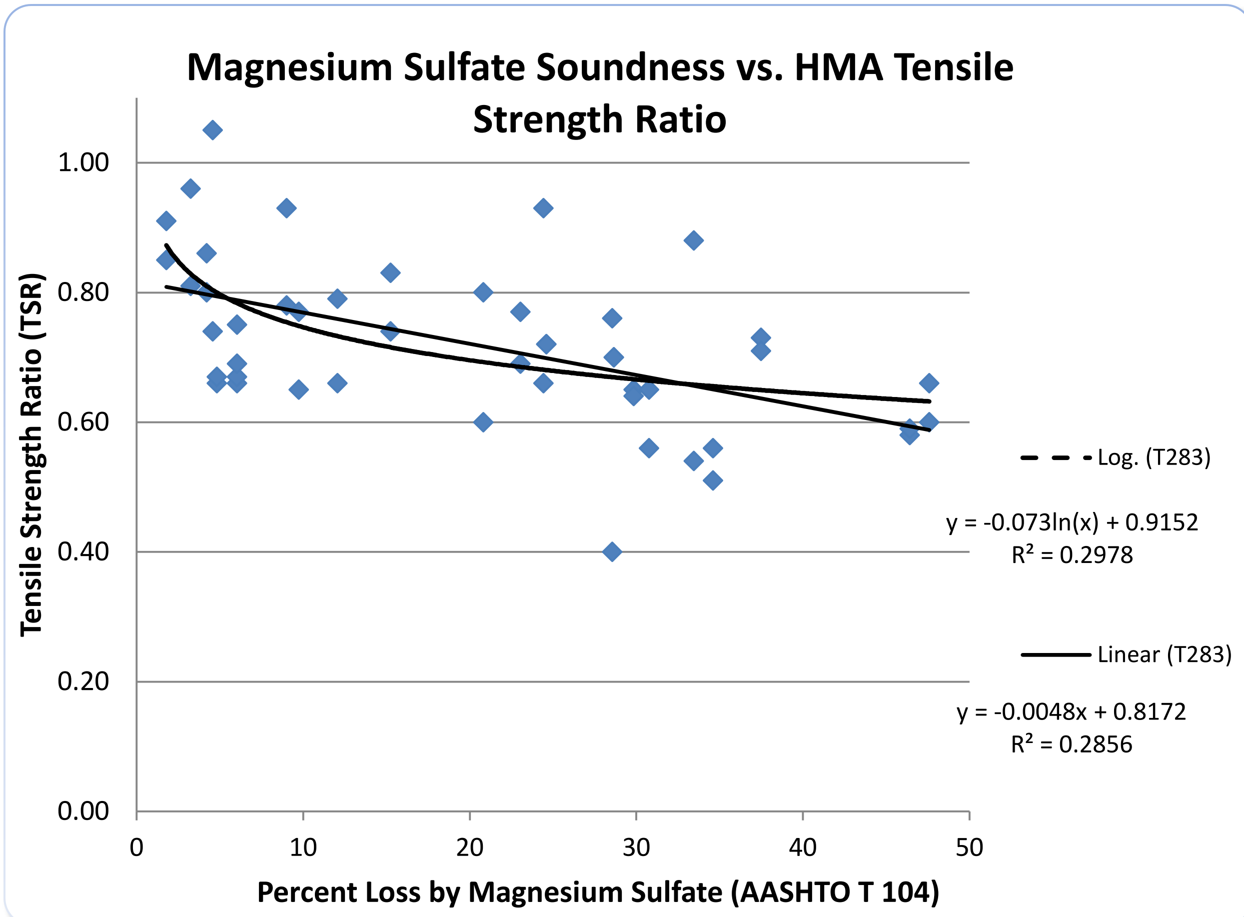

Figure 14. Relationship of Magnesium Sulfate Soundness and HMA Tensile Strength Ratio

The sodium sulfate soundness method was determined to provide the best prediction of rut depth at 10,000 cycles. However, the precision of the sodium sulfate method was questionable, making this relationship less valuable. Figure 15 provides an illustration of the relationship of these parameters. The polynomial relationship was the best non-linear regression technique, but appears very near linear. As rutting performance decreases, soundness performance also decreases, indicating a simultaneous loss of performance in the aggregate and HMA materials. The low associated R^2 values suggested a weak relationship.

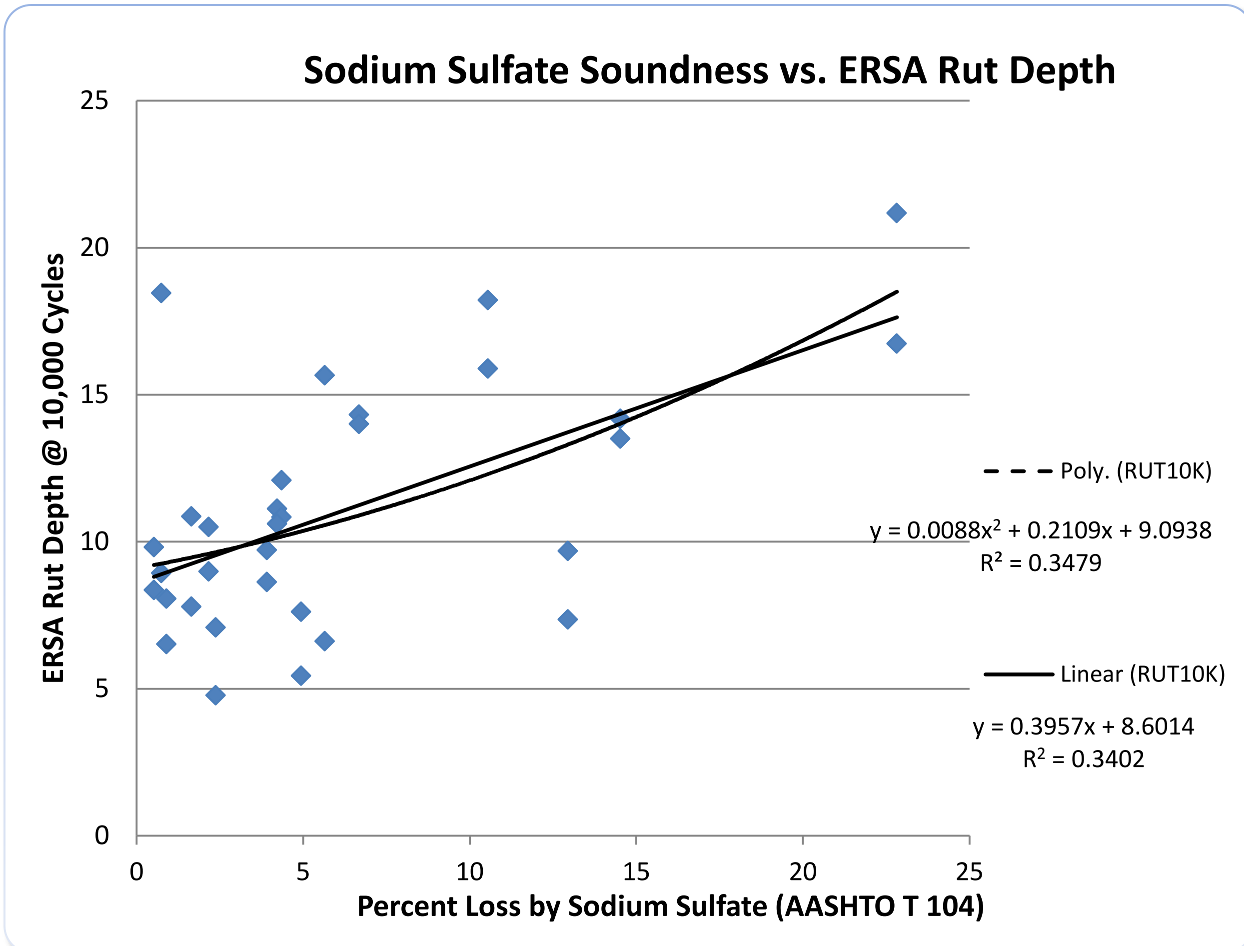

Figure 15. Relationship of Sodium Sulfate Soundness and ERSA Rut Depth

Finally, the relationship of ERSA Stripping Inflection Point and percent loss by the Micro-Deval method was further investigated, and is illustrated in Figure 16. As Micro-Deval percent loss increased, stripping performance decreased. This reasonable relationship was nearly linear, but heavily affected by the small number of samples that did not strip. The elimination of these highly influential data points would change the relationship significantly (reducing the R^2 value), but would also discredit an important feature of the dataset. In order to numerically account for these data points, an arbitrary value of 40,000 cycles was assigned to samples that did not strip.

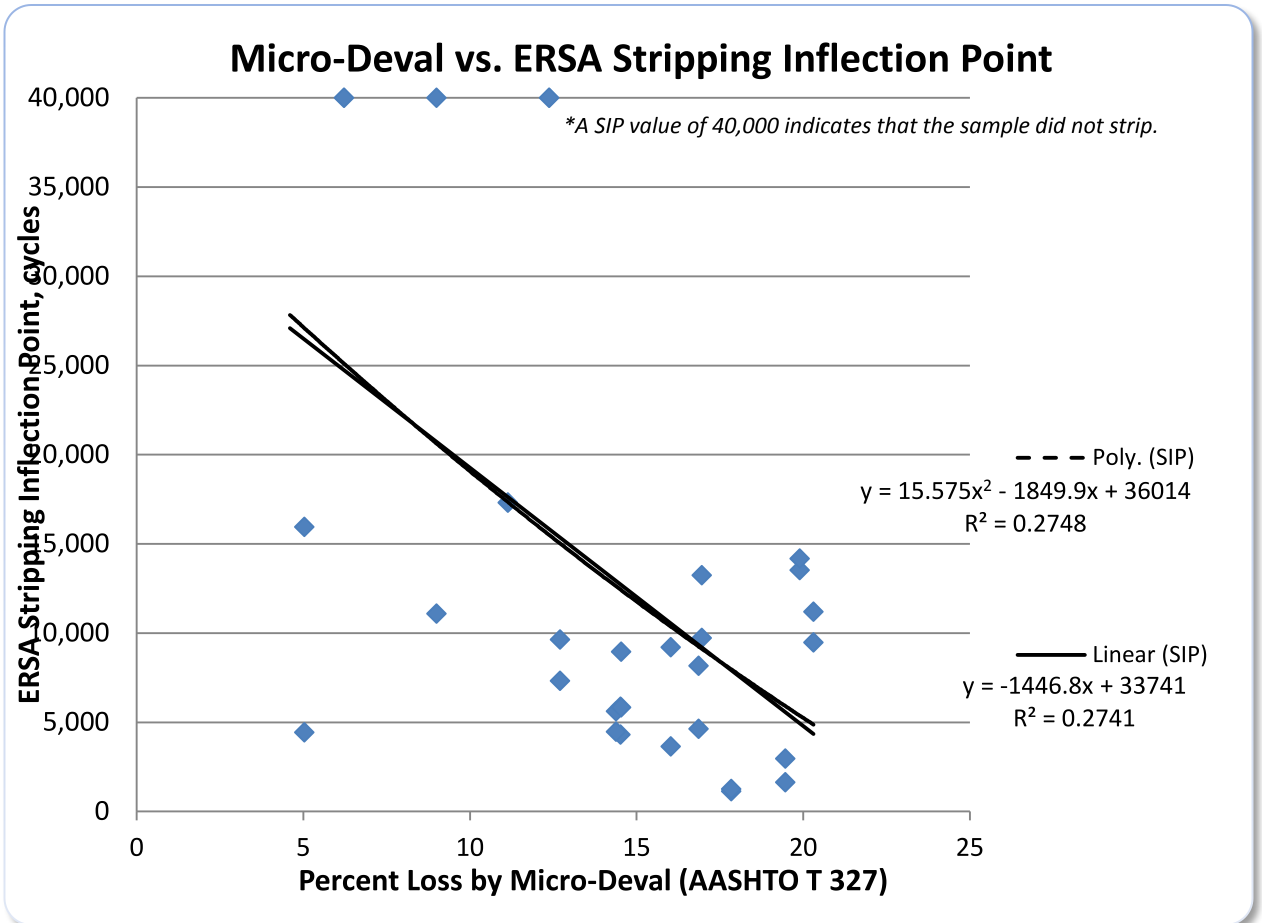

Figure 16. Relationship of Micro-Deval Abrasion and ERSA Stripping Inflection Point

PCC Performance

Finally, the same regression techniques were used to compare aggregate rank and performance to concrete mixture performance measures. First, the concrete performance characteristics were used in efforts to predict known aggregate rank. Concrete performance predictor variables included:

- 14-day Strength (STR14)
- 28-day Strength (STR28)
- Strength After 5 Freeze-Thaw Cycles (C5STR)
- Strength After 10 Freeze-Thaw Cycles (C10STR)
- Durability Ratio at 120 Cycles (DR120)
- Durability Ratio at 200 Cycles (DR200)
- Durability Ratio at 300 Cycles (DR300)
- Difference in 28-Day Strength and Strength After 10 Freeze-Thaw Cycles (D28D10C)

From Table 45, it is shown that D28D10C was the single most capable predictor variable (R^2 = 0.34). By using a combination of six predictor variables, the R^2 value was more than doubled, meaning that almost

75 percent of the variability in aggregate rank could be explained by the combined concrete performance parameters. The next best single predictors were DR300 (R^2 = 0.22) and C10STR (R^2 = 0.20). Clearly, no single variable proved to be adequate for soundness prediction rankings.

Table 45. Regression Summary Using Concrete Performance to Predict Aggregate Rank

Regression Method	Significant Factors	Model	R^2
Stepwise	D28D10C	RANK = 3.774 + 0.004(D28D10C)	0.34
Backward	C5STR, C10STR, DR300	RANK = 15.692 + 0.003(C5STR) – 0.005(C10STR) – 0.041(DR300)	0.62
Best R^2	STR14, C5STR, C10STR, DR120, DR200, DR300	RANK = 28.315 + 0.004(STR14) + 0.003(C5STR) – 0.008(C10STR) – 0.290(DR120) + 0.150(DR200) – 0.029(DR300)	0.74
Most significant single factor	D28D10C	RANK = 3.774 + 0.004(D28D10C)	0.34

Next, aggregate soundness properties were set as the independent variables in an attempt to describe various concrete performance properties. Only the more significant relationships are included in this section. Summary regression data is given in Tables 46 through 51.

Table 46. Regression Summary Using Aggregate Soundness to Predict 14-day Compressive Strength

Regression Method	Significant Factors	Model	R^2
Stepwise	MDV	STR14 = 5902.6 – 86.1(MDV)	0.40
Backward	MDV	STR14 = 5902.6 – 86.1(MDV)	0.40
Best R^2	SSPL, MDV	STR14 = 5860.7 – 24.0(SSPL) – 72.6(MDV)	0.45
Most significant single factor	MDV	STR14 = 5902.6 – 86.1(MDV)	0.40

Table 47. Regression Summary Using Aggregate Soundness to Predict Compressive Strength after 10 Freeze-Thaw Cycles

Regression Method	Significant Factors	Model	R^2
Stepwise	DFRZ, MDV	C10STR = 6670.3 – 30.6(DFRZ) – 64.1(MDV)	0.44
Backward	MDV, DFRZ	C10STR = 6670.3 – 64.1(MDV) – 30.6(DFRZ)	0.44
Best R^2	MDV, FT103, DFRZ, ABS	C10STR = 6640.1 – 74.5(MDV) – 15.8(FT103)-38.9(DFRZ) + 282.0(ABS)	0.50
Most significant single factor	DFRZ	C10STR = 5871.6 – 39.4(DFRZ)	0.29

Table 48. Regression Summary Using Aggregate Soundness to Predict Durability Ratio at 120 Days

Regression Method	Significant Factors	Model	R^2
Stepwise	MDV, DFRZ	DR120 = 99.77 – 0.59(MDV) + 0.27(DFRZ)	0.42
Backward	MSPL, MDV	DR120 = 100.55 + 0.21(MSPL) – 0.70(MDV)	0.40
Best R^2	MSPL, MDV, DFRZ, ABS	DR120 = 102.18 + 0.28(MSPL) – 0.59(MDV) + 0.27(DFRZ) – 4.36(ABS)	0.58
Most significant single factor	DFRZ	DR120 = 92.39 = 0.19(DFRZ)	0.15

Table 49. Regression Summary Using Aggregate Soundness to Predict Durability Ratio at 200 Days

Regression Method	Significant Factors	Model	R^2
Stepwise	MDV, FT103	DR200 = 92.84 – 1.15(MDV) + 0.36(FT103)	0.48
Backward	MDV, FT103	DR200 = 92.84 – 1.15(MDV) + 0.36(FT103)	0.48
Best R^2	MSPL, MDV, FT103	DR200 = 93.54 + 0.19(MSPL) – 1.35(MDV) + 0.24(FT103)	0.52
Most significant single factor	MDV	DR200 = 95.40 – 0.96(MDV)	0.24

Table 50. Regression Summary Using Aggregate Soundness to Predict Durability Ratio at 300 Days

Regression Method	Significant Factors	Model	R^2
Stepwise	MSPL, SSPL, DFRZ, ABS	DR300 = 111.87 – 3.13(MSPL) + 7.16(SSPL) + 2.86(DFRZ) – 39.11(ABS)	0.65
Backward	MSPL, SSPL, DFRZ, ABS	DR300 = 111.87 – 3.13(MSPL) + 7.16(SSPL) + 2.86(DFRZ) – 39.11(ABS)	0.65
Best R^2	MSPL, SSPL, DFRZ, ABS	DR300 = 111.87 – 3.13(MSPL) + 7.16(SSPL) + 2.86(DFRZ) – 39.11(ABS)	0.65
Most significant single factor	MSPL	DR300 = 83.30 – 1.26(MSPL)	0.21

Table 51. Regression Summary Using Aggregate Soundness to Predict the Difference in 28-day Strength and Strength After 10 Freeze-Thaw Cycles

Regression Method	Significant Factors	Model	R^2
Stepwise	SSPL, DFRZ	D28D10C = -88.77 – 18.59(SSPL) + 30.51(DFRZ)	0.60
Backward	DFRZ	D28D10C = -128.56 + 24.58(DFRZ)	0.52
Best R^2	SSPL, MSPL, DFRZ	D28D10C = -122.40 – 26.08(SSPL) + 7.13(MSPL) + 25.43(DFRZ)	0.62
Most significant single factor	DFRZ	D28D10C = -128.56 + 24.58(DFRZ)	0.52

Further evaluations were performed using alternative regression techniques to correlate the top predictive concrete performance test methods with aggregate soundness measures. Percent loss by the magnesium sulfate method, the most advantageous measure of aggregate soundness, was the most influential soundness parameter relating to durability ratio after 300 freeze-thaw cycles. The linear and polynomial relationships are shown in Figure 17. Although the polynomial fit increased the R^2 value to more than twice that of the linear relationship, it is again evident that a relatively small number of data points significantly influenced the dataset. Also the relationship is not consistent. One would expect the durability ratio to decrease with increased magnesium sulfate percent loss; however, this is not clearly depicted in the graph.

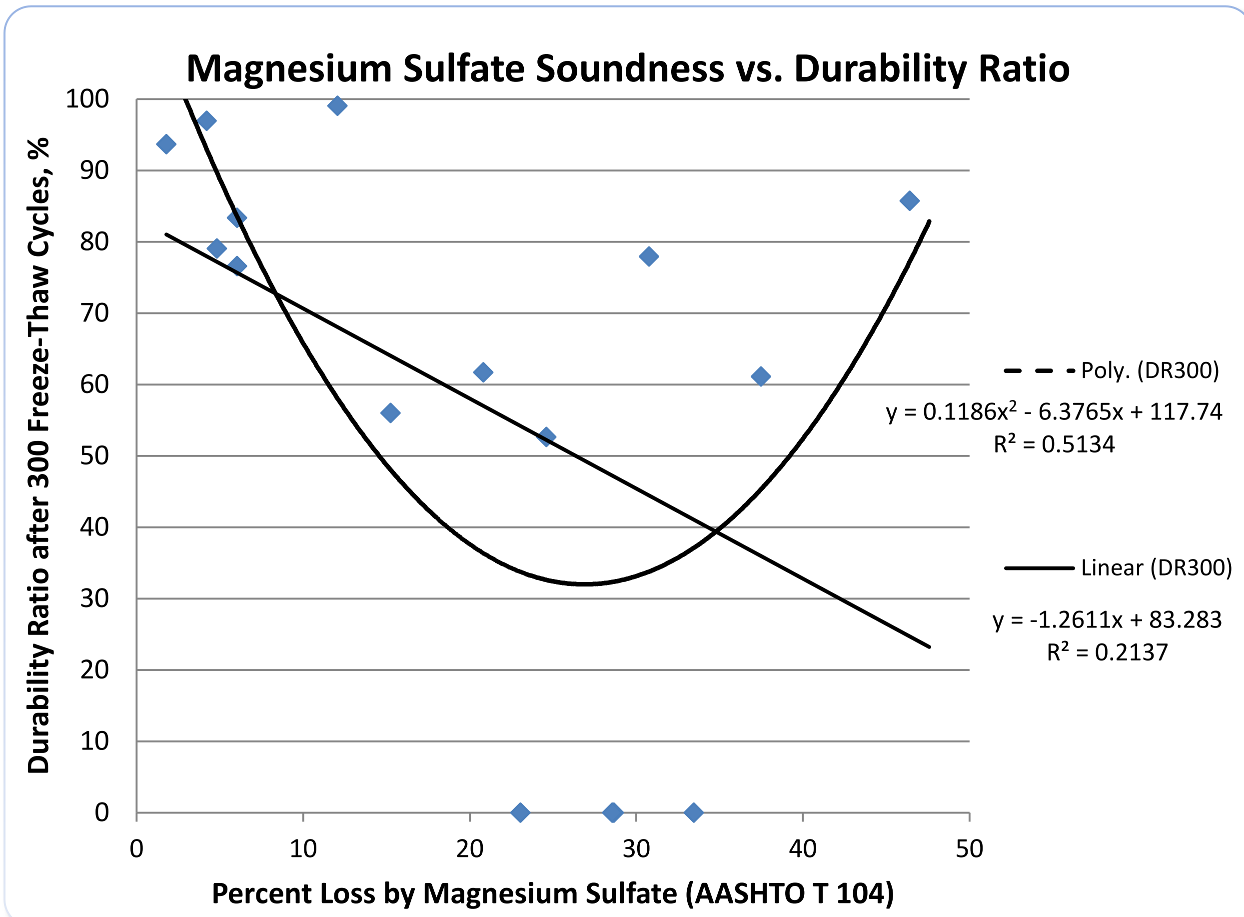

Figure 17. Relationship of Magnesium Sulfate Soundness and Durability Ratio

Next, the strongest relationship of aggregate and concrete mixture performance was further investigated. Figure 18 displays the relationship of the percent loss of aggregate by the Deep Freeze method to the difference in 28-day concrete compressive strength and strength after 10 freeze-thaw cycles. Although the polynomial relationship was identified as the most significant non-linear regression method, this trend line appeared almost identical to the linear relationship. In this graph, it is evident that as aggregate quality decreased (i.e., percent loss by the deep freeze method increased), the loss of concrete compressive strength increased. Specimens displaying a negative loss of strength are considered as having no loss of strength, indicating that the concrete mixture was not affected by the freeze-thaw conditioning cycles.

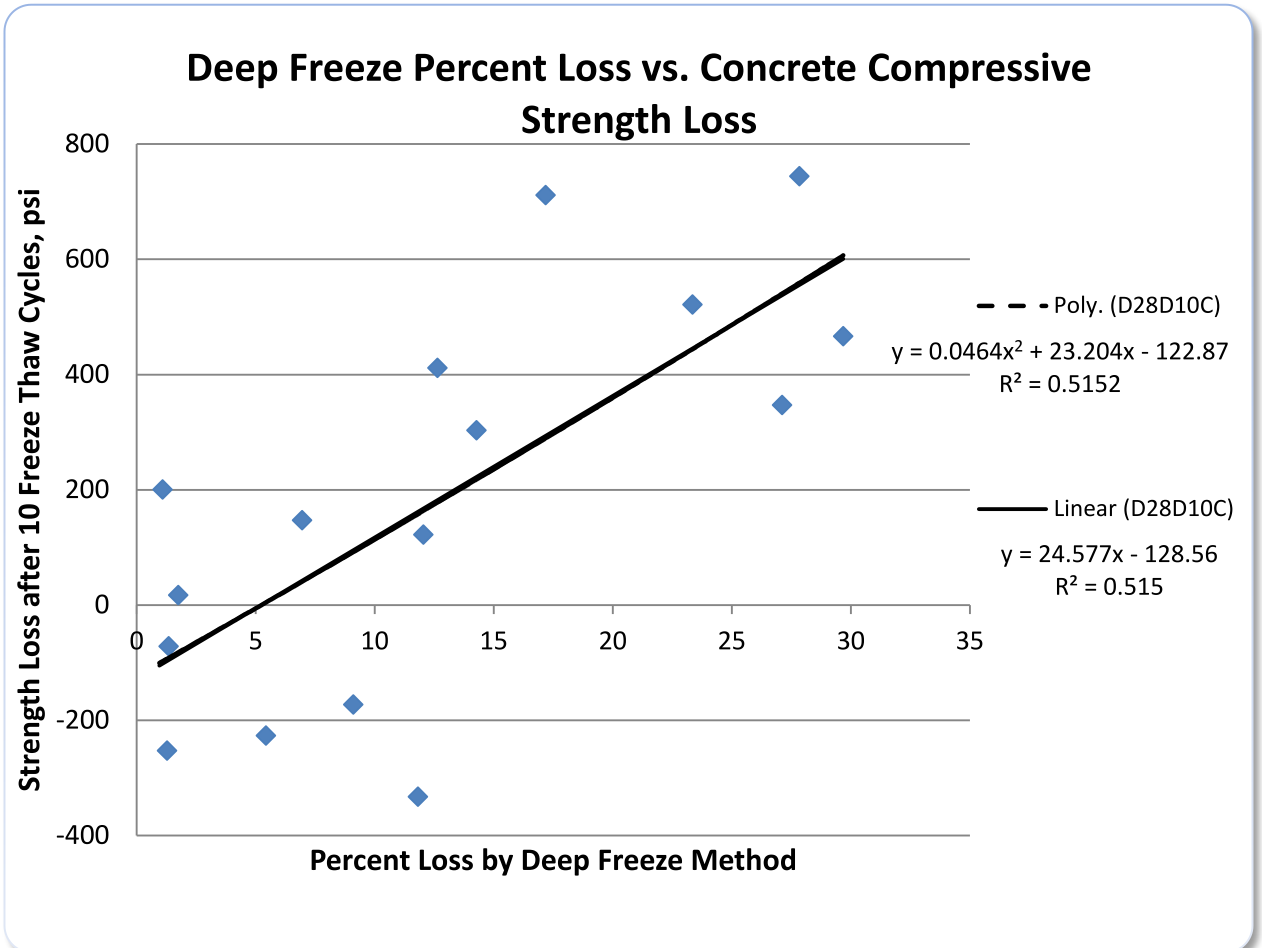

Figure 18. Relationship of Deep Freeze Soundness and Loss of Strength After 10 Freeze-Thaw Cycles

Although many of the concrete mixture performance parameters were able to discern between varying levels of aggregate quality, these relationships were certainly not capable for predictive purposes. The most significant relationships were noted for the magnesium sulfate soundness test and durability ratio after 300 cycles, and aggregate percent loss by the deep freeze method and concrete compressive strength loss after 10 freeze-thaw cycles. As was demonstrated by the asphalt performance measures, it is suspected that the confounding effects of other mixture properties interfered with relationships based solely on aggregate properties.

<u>Practicality of Test Methods</u>

In order to choose a test method that is most advantageous for use in qualifying or disqualifying aggregate sources based on soundness characteristics, several things should be considered. The method must be able to reasonably predict the performance of the aggregate, it should be able to distinguish between varying levels of aggregate quality, and should be practical for incorporation into current testing procedures.

Based on the discussions of variability, accuracy, and discrimination, the T85 and vacuum saturation measures of absorption were consistently mentioned as capable methods. These methods are relatively simple to perform, and are calculated from data recorded during the specific gravity determination. The vacuum saturation method for determining absorption capacity did not present any considerable advantage over the traditional T 85 procedure; therefore, the T 85 method should be considered as an appropriate measure of absorption capacity. This method would be simple to incorporate into current specifications, as T 85 is already a standard test method referenced in the Gold Book. Absorption capacity is intuitively related to aggregate soundness properties because the greater the capacity of the aggregate to take on water, the greater its likelihood to retain the moisture that, when frozen, would expand and apply excessive stresses to the aggregate structure.

The sodium sulfate soundness test was able to rank the aggregate sources fairly accurately, but exhibited excessive variability. While ranking is important, it represents a relative comparison, and may not be accurately reflected by finite specification limits. Additionally, the high level of variability greatly increases the chances of incorrectly qualifying or disqualifying an aggregate source, or would require a much larger sample size in order to provide confidence in test results. Because the test is relatively time consuming and requires tedious attention, greatly increasing the number of samples required would not be a readily accepted alternative. An alternative test method would likely provide greater advantages.

The magnesium sulfate soundness test was a fairly good performer in terms of variability, and was a good performer in terms of accuracy and discrimination. This test method is lengthy and difficult, requiring careful temperature control. The salt required for the solution is considerably more expensive than that needed for the sodium sulfate version of the test method, but does provide better discrimination and lower variability. Another advantage of this method is that it is intended specifically to provide a measure of soundness (rather than toughness), and simulates the environmental cycling that occurs in the field. If AASHTO T 104 is to be included in the standard specification, the magnesium sulfate solution should be used, and the specification limits increased to reflect aggregate quality limits.

In terms of variability, the Micro-Deval method was a top performer, but was not successful at ranking aggregate soundness performance. This is not unexpected since the Micro-Deval test is performed under conditions that mimic the abrading and grinding action that could occur during production and construction, but does not apply temperature as a conditioning factor. Thus, this test method is more of a toughness test than a soundness test, and should not be used as the sole qualifier for aggregate soundness.

The aggregate freeze thaw tests (AASHTO T 103 and the deep freeze method) simulate the freezing and thawing action that creates soundness issues in the field, but in an accelerated form. In theory, these methods should provide the most accurate measure of an aggregate's ability to withstand naturallyoccurring temperature cycles because the laboratory process most closely mimics field conditions. However, the variability associated with these methods did not generate a great deal of confidence. The simplified deep-freeze method actually provided a more accurate prediction of performance than AASHTO T 103, and did not require expensive laboratory equipment. While the deep freeze method did not offer as many advantages as some of the other methods, it could provide a viable alternative to AASHTO T 103, while subjecting the aggregate to a more realistic freeze-thaw conditioning process.

7. Conclusions and Recommendations

In this study, eight aggregate sources were used to assess the ability of various aggregate soundness measures to quantify the performance of carbonate aggregate sources in Arkansas. In addition to aggregate soundness tests, laboratory mixture performance tests were employed to assess the performance of each aggregate when used as the primary aggregate component in asphalt and concrete paving mixtures.

In terms of aggregate soundness, several test methods were used to assess the soundness properties of each aggregate source. These methods included:

- Sodium Sulfate Soundness (AASHTO T 104)
- Magnesium Sulfate Soundness (AASHTO T 104)
- Micro-Deval Abrasion (AASHTO T 327)
- Aggregate Freeze-Thaw (AASHTO T 103)
- Aggregate Freeze-Thaw (Deep Freeze Method)
- Coarse Aggregate Absorption Capacity (AASHTO T 85)
- Absorption Capacity (Vacuum Saturation Method)

Overall, the most advantageous parameter was percent loss by the magnesium sulfate soundness method, according to AASHTO T 104. Although this method was not identified as having the least overall variability, a very low percentage of pure error was associated with this method, unlike its sodium sulfate counterpart. It was identified as one of the most capable methods for ranking aggregate soundness performance (based on known historical performance) and it was also judged as very capable in distinguishing between varying levels of actual aggregate performance. Aggregate absorption capacity was informative, and data comparisons supported the claim that aggregates having an absorption capacity greater than 2 percent were more prone to distress resulting from soundness problems. Absorption capacity by the vacuum saturation method was no more advantageous than absorption capacity by AASHTO T 85. Therefore the currently specified method, T 85, is deemed adequate for these determinations.

The sodium sulfate test displayed an excessive level of variability, casting doubt on its ability to accurately qualify or disqualify aggregate sources. Even though the rankings provided by this method were very reasonable, the variability issues significantly decreased its reliability.

The Micro-Deval method was determined to have good repeatability, but was unable to consistently rank the aggregates in terms of soundness performance. It was believed that the mechanism used to abrade the samples may provide accurate predictions of physical aggregate breakdown, but did not adequately represent the aggregates' ability to withstand environmental freeze-thaw conditioning.

Two methods were used to test the freeze-thaw performance of each aggregate using actual freeze-thaw cycles. The first was performed using an automatically controlled freeze-thaw chamber according to AASHTO T 103, and the other was a simplified version of the test method using fewer freeze-thaw cycles in a deep freezer. The results of each method were fair, with the deep freeze method displaying slightly better performance in terms of accuracy and consistency. This method should be considered as an alternative for providing additional aggregate performance information.

After soundness testing, each aggregate source was then used in HMA mix designs, and tested for rutting, stripping, and durability performance using the following methods:

- Evaluator of Rutting and Stripping in Asphalt (ERSA)
- Resistance of Compacted Hot Mix Asphalt to Moisture Induced Damage (AASHTO T 283)
- Cantabro Loss (TxDOT Method, TEX-245-F)

The known aggregate rankings were most closely reflected by the stripping inflection point as determined from the ERSA test. TSR (by AASHTO T 283) was also an adequate indicator; however, these methods were not always capable of adequately discriminating between varying levels of aggregate quality. Rut depth at 10,000 cycles was able to discern between different levels of HMA mixture performance, but was only marginally capable of correctly ranking the aggregate sources in terms of performance. Because many features of an asphalt mixture affect rutting and stripping performance (in addition to aggregate soundness performance), this is not an unreasonable result. The addition of a freeze-thaw cycle in the ERSA testing regimen did not significantly affect results.

AASHTO T 283 was used to measure TSR, dry tensile strength, and wet tensile strength. TSR was most adept at correctly ranking the aggregate sources. Additional testing was performed for this method using specimens cored to a 4-inch diameter, exposing aggregate faces and accelerating the potential for distress. Although the wet and dry tensile strengths were significantly affected by the coring process, the resulting TSR values were not.

The Cantabro loss test was reasonably able to detect the good and bad performers, but had difficulty discerning the marginal performers. Because this test was intended to be more of a physical durability test than a measure of the ability to withstand environmental conditioning, this result was consistent with expectations.

Concrete mixture performance was investigated with a particular focus on test methods relating to environmental conditioning. The test methods used included:

- Resistance of Concrete to Rapid Freezing and Thawing (ASTM C 666)
- Durability Ratio of Concrete Specimens by Resonant Frequency (ASTM C 215)
- Compressive Strength of Concrete Using Freeze-Thaw Cycles

The most capable measures of concrete performance for predicting aggregate rank were durability ratio after 300 cycles and the loss of compressive strength after 10 freeze-thaw cycles (i.e., the difference in 28-day compressive strength and strength after 10 freeze-thaw cycles). Unfortunately, these measures were also judged as incapable of significantly distinguishing between aggregate quality levels, but did provide a link between aggregate performance and concrete mixture performance. Compressive strength values were able to detect differences in aggregate type, but did not necessarily provide an indication of the soundness performance of the aggregates. It is likely that the many features of the concrete mix design process confounded the ability to detect the effects of individual aggregates.

Based on the results of this project, it is recommended that the magnesium sulfate soundness test (AASHTO T 104) be specified in place of the current sodium sulfate soundness requirement, with a recommended maximum percent loss of 18 percent. It is also recommended that carbonate aggregates possessing an absorption capacity of more than 2.0 percent be subjected to further testing by the Aggregate Freeze-Thaw by Deep Freeze method, allowing a maximum of 15 percent loss. A draft of this method is given in Appendix A. By incorporating these changes to the Standard Specification, an increased level of testing precision may be achieved, leading to greater confidence in decisions regarding the qualification of aggregate sources.

References

American Association of State Highway and Transportation Officials. (2009). *Standard Specifications for Transportation Materials and Methods of Sampling and Testing.* Twenty-ninth Edition, Washington, D.C.

Arkansas State Highway and Transportation Department (2003). *Standard Specifications for Highway Construction.* Little Rock, Arkansas.

Arkansas State Highway and Transportation Department (2009), *Manual of Field Sampling and Testing Procedures.* updated July 2009, Materials Division, Little Rock, Arkansas.

ASTM. (2010). *Annual Book of ASTM Standards*. West Conshohocken, Pennsylvania.

Azari, H. (2010). Precision Estimates of AASHTO T 283: Resistance of Compacted Hot Mix Asphalt (HMA) to Moisture-Induced Damage. *Publication NCHRP Project 9-26A*. Transportation Research Board of the National Academies, Washington, D.C.

Azari, H. (2011). Precision Statements for AASHTO Standard Methods of Test T 148, T 265, T 267, and T 283. *Publication Research Results Digest 351*. Transportation Research Board of the National Academies, Washington, D.C.

Barksdale, R.D. (1991). *The Aggregate Handbook*, National Stone Association, Washington, D.C.

Bjarnason, G., P. Petursson, and S. Erlingsson. (2002). *Quality Assessment of Aggregates for Road Construction: Fragmentation, Weathering, and Abrasion*. Presented at Ráðstefna norrænna steinefnaframleiðenda, Public Roads Administration, Reykjavik, Iceland, September 12, 2002.

Bloem, D.L. (1966). Soundness and Deleterious Substances, Significance of Tests and Properties of Concrete and Concrete-Making Materials. *ASTM STP 169A,* ASTM International, West Conshohocken, PA, 497-512.

Brandes, H. G., and C. E. Robinson. (2006). Correlation of Aggregate Test Parameters to Hot Mix Asphalt Pavement Performance in Hawaii. *Journal of Transportation Engineering*, Vol. 132, No. 1, 86-95.

Cooley, Jr., L.A. and R.S. James. (2004). Use of Micro-Deval Test to Evaluate the Toughness of Coarse Aggregates to be Used in Hot Mix Asphalt*.* 2004 Transportation Systems Workshop, Ft. Lauderdale, Florida.

Cuelho, E., Mokwa, R., and K. Obert. (2007). Comparative Analysis of Coarse Surfacing Aggregate Using Micro-Deval, L.A. Abrasion and Sodium Sulfate Soundness Tests. *Project Summary Report 8117-27*, Western Transportation Institute, Montana Department of Transportation.

Cuelho, E., Mokwa, R., Obert, K., and A. Miller. (2008) Comparative Analysis of Micro-Deval, L.A. Abrasion and Sulfate Soundness Tests. In *TRB 2008 Annual Meeting CD-ROM*, Transportation Research Board of the National Academies, Washington, D.C.

Federal Highway Administration. (2003). *Distress Identification Manual for the Long-Term Pavement Performance Program.* Publication No. FHWA-RD 03-031.

Federal Highway Administration, U.S. Department of Transportation. *Asphalt Pavement Technology.* Washington, D.C. Retrieved Dec. 1, 2010, from http://www.fhwa.dot.gov/pavement/asphalt/labs/mixtures/hamburg.cfm

Hall, K.D., and S.G. Williams. (1998). Acquisition and Evaluation of Hamburg Wheel-Tracking Device. *Publication MBTC FR-1044*. Department of Civil Engineering, University of Arkansas, Fayetteville, Arkansas.

Hamilton, R.D., Smith, R.E., and G.B. Sherman. (1971). Factors Influencing the Durability of Aggregates. *Research Report 633476*, State of California, Division of Highways, Materials, and Research Department.

Hooton, R.D., and Rogers, C.A. (1989). Evaluating Rapid Test Methods for Detecting Alkali-Reactive Aggregates. *Proceedings of the 8th International Conference on Alkali-Aggregate Reaction in Concrete, Kyoto, Japan*. Edited by K. Okada, S. Nishibayashi, and M. Kawamura, Engineering Materials Report 92. Elsevier, 439-444.

Alden, Andrew. Geologic map of Arkansas. (1974). Created from the U.S. Geological Survey's Geologic Map of the United States, Retrieved on April 24, 2009 from http://geology.about.com/library/bl/maps/n_statemap_AR2000.htm.

Janoo, V.C., and C. Korhonen. (1999). Performance Testing of Hot Mix Asphalt Aggregates. *Special Report 99-20*. US Army Corps of Engineers.

Kiggundu, B. M. and F. L. Roberts. (1988). Stripping in HMA Mixtures: State-of-the-Art and Critical Review of Test Methods. *Publication NCAT Report No. 88-2*. National Center for Asphalt Technology of Auburn University, Auburn, Alabama.

Kline, S.W., Phiukao, W., Griffin, M., & Miller, J.W. (2004), Evaluation of the Sodium Sulfate Soundness Test for Qualifying Dolomites of Northern Arkansas for Construction Aggregate: Proceedings of the 40th Forum on the Geology of Industrial Minerals, Bloomington, Indiana. *Indiana Geological Survey Occasional Paper 67*, 115-128.

Martin, A. E., Button, J., Estakhri, C, and Glover, C. (2007). Optimizing the Design of Permeable Friction Courses (PFC). *Project Summary 0-5262*, Texas Transportation Institute (TTI), Texas A&M University System, College Station, Texas.

Meininger, R.C. (2002). Validity of the Sulfate Soundness Test. Retrieved on April 26, 2010 from http://rockproducts.com/mag/rock_validity_sulfate_soundness/index.html.

Nokken, M.R., R.D. Hooton and C.A. Rogers. (2004). Measured Internal Temperatures in Concrete Exposed to Outdoor Cyclic Freezing. *Cement, Concrete and Aggregates*, Vol. 26, No. 1, 26-32.

Powers, T.C., Copeland, L.E., Hayes, J.C. and H. M. Mann. (1955). Permeability of Portland Cement Paste. *Journal of the American Concrete Institute*, 51, 285-298.

Rangaraju, P. R., and J. Edlinski. (2008). Comparative Evaluation of Micro-Deval Abrasion Test with Other Toughness/Abrasion Resistance and Soundness Tests. *Journal of Materials in Civil Engineering*, Vol. 20, No. 5, 343-351.

Senior, S. A., and C. A. Rogers. (1991). Laboratory Tests for Predicting Coarse Aggregate Performance in Ontario. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1301, Transportation Research Board of the National Academies, Washington, D.C., 97-106.

Stark, D. (1976). Characteristics and Utilization of Coarse Aggregates Associated with D-Cracking, *Bulletin RD047.01P*, Portland Cement Association, Skokie, Illinois.

Texas Department of Transportation. (2005). Test Procedure for Cantabro Loss. *TxDOT Designation Tex-245-F*, October.

Verbeck, G. and R. Landgren. Influence of Physical Characteristics of Aggregates on Frost Resistance of Concrete. *PCA Reseach Development Bulletin 126*.

Washington State Department of Transportation. (2010). *WSDOT Pavement Guide.* Retrieved on December 1, 2010 from http://training.ce.washington.edu/WSDOT/

Weyers, R. E., G.S. Williamson, D. W. Mokarem, D. S. Lane, and P. D. Cady. (2005). Testing Methods to Determine Long Term Durability of Wisconsin Aggregate Resources. *Publication No. WHRP 06-07,* Virginia Polytechnic Institute and State University, Wisconsin Department of Transportation.

Williams, S.G. (2001). *Development of Wheel-Tracking Test Method and Performance Criteria for Asphalt Pavement*. Doctoral Dissertation, University of Arkansas, Fayetteville, Arkansas.

Williamson, G. S., Weyers, R.E., Mckarem, D. W., Lane, D.S., and D.D. Reid. (2007). Vacuum Saturated Absorption As Aggregate Durability Indicator. *ACI Materials Journal*, Technical Paper, Vol. 104, No. 3, 307 - 312.

Wu, Y., Parker, F., and K. Kandhal. (1998). Aggregate Toughness/Abrasion Resistance and Durability/Soundness Tests Related to Asphalt Concrete Performance in Pavements. *NCAT Report No. 98-4*, National Center for Asphalt Technology, Auburn, Alabama.

Yildirim, Y., Jayawickrama, P. W., Hossain, M. S., Alhabshi, A., Yildirim, C., Smit, A., and D. Little. (2006). Hamburg Wheel-Tracking Database Analysis. *FHWA/TX-05/0-1707-7*, Texas Transportation Institute, The Texas A&M University System, College Station, Texas.

Yildirim, Y. and K. H. Stokoe II. (2006). Analysis of Hamburg Wheel Tracking Device Results in Relation to Field Performance. *Project Summary Report 0-4185-S*, Center for Transportation Research, The University of Texas, Austin, Texas.

APPENDIX A

Draft Test Procedure for

Aggregate Freeze-Thaw by Deep Freeze Method

AHTD Test Method XXX-XX
Effective Date: July 2012

1. SCOPE

1.1. This test method provides a numerical measure of an aggregate's resistance to breakdown due to freeze-thaw conditioning using a deep freeze.

2. REFRENCED DOCUMENTS

2.1. American Assocation of State Highway and Transportation Officials (AASHTO) Standards:

2.1.1. T 103 Soundness of Aggregates by Freezing and Thawing

3. SIGNIFICANCE AND USE

3.1. This test method is used for determining the percent loss of aggregate sources that have been determined to be "high risk" according to absorption testing (i.e., greater than 2 percent absorption capacity) and shall be performed in addition to the magnesium sulfate soundness test.

4. APPARATUS

4.1. Freezing equipment – a residential grade freezer unit capable of maintaining temperatures at least as low as -18°C (0°F).

4.2. Sample containers – The sample containers shall be of plastic, rubber, or other suitable materials and shall have close-fitting lids. The containers shall be of sufficient size for containing the entire specimen submerged in solution throughout the duration of the testing procedure.

4.3. Sieves – The sieves used shall meet the requirements of M 92.

4.4. Balance – The balance shall have sufficient capacity, be readable to 0.1 percent of the sample mass, or better, and conform to the requirements of M 231.

4.5. Drying Oven – The drying oven shall provide a free circulation of air through the oven and shall be capable of maintaining a temperature of 110° ± 5°C (230° ± 9°F).

---

5. PROCEDURE

5.1. Obtain a representative sample of the aggregate in accordance with AASHTO T 2 and T 248 in order to obtain the necessary quantity of aggregate.

5.2. Prepare the aggregate sample in accordance with AASHTO T 103, sections 4 and 5.

5.3. Each sample fraction shall be placed in a separate freeze-thaw container and cover until completely immersed with a 0.5 percent isopropyl alcohol and water solution.

5.4. Allow the immersed sample to soak at room temperature for a period of 24 ± 4 hours.

5.5. Place each sample container (i.e., each aggregate fraction) in the freezer for a period of 24 ± 4 hours. Remove the sample from the freezer and allow to stand at room temperature for a period of 24 ± 4 hours. This constitutes one freeze-thaw cycle.

5.6. Repeat section 5.5 until 10 freeze-thaw cycles have been completed. If at any time the test must be interrupted, store the samples in the thawed condition until testing can be resumed.

5.7. Perform a quantitative examination as described in AASHTO T 103, section 8.

---

6. REPORT

6.1. The report shall include the following data:

6.1.1. Mass of each fraction of each sample before test.

6.1.2. Actual loss of each fraction of the sample expressed as a percentage of the original mass of the fraction.

6.1.3. Weighted average calculated from the percentage of loss for each fraction, based on the representative grading of the sample. In these calculations, sizes finer than the 300-μm (No. 50) sieve shall be assumed to have zero-percent loss.