

**Evaluation of the Alabama Accountability Act:  
Academic Achievement Test Outcomes of Scholarship  
Recipients 2014-2015**

---

**The Institute for Social Science Research**

**The University of Alabama**

Joan M. Barth  
Garrett Quenneville

September 1, 2016

## Table of Contents

Executive Summary .....	ii
List of Abbreviations .....	iii
Acknowledgements.....	iii
Introduction.....	1
Scholarship Recipient Testing Requirements.....	1
Evaluation Reporting Requirements .....	1
Alabama Mandated State Testing in Public Schools 2014-2015 Academic Year .....	1
Timeline.....	2
Method .....	2
Challenges.....	2
Demographic Description of Scholarship Recipients .....	4
Scholarship Recipients with Test Data .....	4
Demographic Information for Scholarship Recipients Included in the Evaluation .....	6
Objective 1: Describe the Academic Achievement of Students in the Scholarship Program .....	6
Stanford Achievement Test 10 .....	7
Iowa Test of Basic Skills .....	9
Terra Nova .....	9
Summary for Stanford 10, Iowa Test of Basic Skills, and Terra Nova .....	10
ACT Practice Test .....	11
ACT Aspire .....	11
Summary of ACT Practice Test and ACT Aspire Test .....	14
Objective 1 Conclusions .....	14
Objective 2: Compare Scholarship Recipients to Alabama Public School Students .....	14
Objective 2 Conclusions .....	17
General Conclusions .....	17
Improving the Evaluation .....	17
Glossary of Terms.....	19
Links to Online Resources .....	20

## Executive Summary

This report fulfills the evaluation requirements of the 2013 Alabama Accountability Act by reporting on the academic achievement of 2014-2015 scholarship recipients.

### **The report focuses on two objectives:**

1. Describe the learning gains of students in the scholarship program.
2. Compare the learning gains of the scholarship recipients to students attending public schools.

Achievement test scores and demographic information were provided by the Scholarship Granting Organizations that administer the student scholarships. Achievement test score information for Alabama public school students was retrieved from the State Department of Education website.

### **Some challenges were encountered in conducting the evaluation:**

- Test scores for the previous year (2013-2014) were not available for most students, and therefore, learning *gains* could not be assessed. A description of the current level of achievement is provided instead.
- The lack of a uniform achievement test among schools constrained the description of the achievement of scholarship recipients and the comparisons that could be made to Alabama public school students.
  - Norm-referenced tests (e.g., the Stanford Achievement Test) and criterion-referenced tests (e.g., ACT Aspire) are based on different standards and cannot be directly compared.
- The test score information available from the Alabama State Department of Education was limited to the percentage of students in proficiency groups based on ACT Aspire, ACT Plan and ACT College Entrance Exam scores, which limited the types of analyses that could be conducted.

The evaluation was based upon test scores from 970 scholarship recipients attending 91 schools and representing 52% of the scholarship recipients in grades for which testing was required.

- 94% were first time scholarship recipients.
- 98% were eligible for free/reduced lunch subsidies.
- 43% were zoned to attend a failing school.
- 65% were Black/African American, 22% were White/Caucasian, and 5% were Hispanic.

*Although this report can show trends for this subsample of scholarship recipients, findings cannot be generalized with confidence to the larger group because the subsample may not be representative of all of the scholarship recipients.*

### **Findings for Objective 1:** Describe the learning gains of students in the scholarship program:

- On norm-referenced tests, scholarship recipients generally performed below the average U.S. student at their grade level.
- On criterion-referenced tests, the majority of scholarship recipients failed to meet benchmark proficiency scores.
- These findings are similar to those of the National Assessment of Educational Progress for students attending public schools in Alabama.

*Continues*

*Executive Summary Continued*

**Findings for Objective 2:** Compare the learning gains of the scholarship recipients to students attending public schools:

- Evaluation of this objective was hampered by the limited number of ACT Aspire and ACT college entrance exam scores available. Comparisons could only be made for grades 6, 7, 10, and 11.
- No cohesive pattern emerged across the different age groups with respect to achievement differences between scholarship recipients and public school students.
- There were very few subject areas in which more than 50% of the students met proficiency standards for either group of students.

**Recommendations for future evaluations:**

- Require a uniform test across schools.
- Require new scholarship recipients to provide the previous year test scores.
- Acquire additional test information from the Department of Education that will allow for more precise comparisons.

## **List of Abbreviations**

AAA	Alabama Accountability Act
AA	African American
AL	Alabama
ALSDE	Alabama State Department of Education
FERPA	Federal Education Rights and Privacy Act
ISSR	Institute for Social Science Research
N	Number of people in a group
n	Number of people in a subgroup
NAEP	National Assessment of Educational Progress
PDF	Portal Document Format
PSAT	The Preliminary SAT/National Merit Scholarship Qualifying Test
SGO	Scholarship Granting Organization

---

## **Acknowledgements**

The authors would like to express their gratitude to Debra McCallum, Director of ISSR for her assistance in developing the report. In addition, Braxton Lowe and William McCallum are recognized for their diligence and hard work in developing the test score data base.

## **Introduction**

The purpose of this report is to fulfill the evaluation component of the 2013 Alabama Accountability Act by reporting on the academic achievement of scholarship recipients in the 2014-2015 academic year. The Alabama Accountability Act (AAA), passed by the legislature in 2013 and amended in 2015, established a scholarship program for low income students to attend public or private schools. Tax-deductible donations for scholarships are managed by Scholarship Granting Organizations (SGOs), which must comply with standards set forth in the Act. The Act places some restrictions on who can receive scholarships based on family income and school zoning. All students receiving scholarships must meet family income eligibility requirements, with priority given to students who are zoned to attend a public school that is failing according to the Alabama State Department of Education (ALSDE) designation; however, students from non-failing public schools may receive scholarships if they meet the income eligibility requirements and additional funds are available. Scholarships are awarded from the SGO to the student to attend a particular school that must meet additional standards set forth in the act. Scholarships may cover all or part of tuition and mandatory fees for one academic year. In 2015, the legislature amended the Act to place limits on the amount that could be awarded depending on the grade level (elementary, middle or high school). The Alabama State Department of Revenue oversees implementation of the Act.

### **Scholarship Recipient Testing Requirements**

The academic accountability standards require the SGOs to ensure that schools accepting scholarship students “annually administer either the state achievement tests or nationally recognized norm-referenced tests that measure learning gains in math and language arts to all students receiving an educational scholarship in grades that require testing under the accountability testing laws of the state for public schools.” The purpose of these tests is to assess the learning gains for scholarship recipients and to provide a means of comparing scholarship recipients to students who attend Alabama public schools.

### **Evaluation Reporting Requirements**

The AAA states that the evaluation shall describe the “learning gains of students receiving educational scholarships and the report shall be aggregated by the grade level, gender, family income level, number of years of participation in the tax credit scholarship program, and race of the student receiving an educational scholarship. The report shall also include, to the extent possible, a comparison of the learning gains of students participating in the tax credit scholarship program to the statewide learning gains of public school students with socioeconomic and educational backgrounds similar to those students participating in the tax credit scholarship program.” The Act also requires that a report be made every two years, starting in 2016. Thus, the first evaluation of the scholarship program had two major objectives: a) to describe the learning gains of students in the scholarship program; and b) to make comparisons between the learning gains of the scholarship recipients and comparable students attending public schools.

### **Alabama Mandated State Testing in Public Schools 2014-2015 Academic Year**

Students attending public schools in Alabama during the 2014-2015 academic year were tested in March and April. Math and reading were assessed with the ACT Aspire for students in grades 3 - 8. ACT Explore, ACT Plan, ACT Plus Writing, and ACT WorkKeys were given for grades 8, 10, 11, and 12, respectively. Although not relevant for this report, students in grades 5 and 7 also took

the ACT science assessment. In addition, for high school students, end-of-course-assessments for Algebra I and English 10 were used to assess achievement, but are not included in this report.

### **Timeline**

On January 11, 2016, the Institute for Social Science Research (ISSR) at The University of Alabama submitted a proposal to the Alabama Department of Revenue to evaluate the scholarship program. On April 12, 2016, a memorandum of understanding was signed between the Alabama Department of Revenue and The University of Alabama. Confidentiality and Disclosure Statement forms from the ISSR staff were completed and returned to the Department of Revenue by April 25<sup>th</sup>. Through the end of April and early May, scholarship recipient test scores and the SGO annual reports were reviewed by ISSR. ISSR communicated with the SGOs, the Department of Revenue, and the ALSDE to collect information to address the evaluation as outlined in the Act. From May through early August, the tasks of the evaluation project were to a) create a data base that listed all scholarship recipients for the 2014-2015 academic year, their demographic information, and test scores, b) research the different standardized tests used by the scholarship recipients' schools, and c) identify the appropriate comparison test score data from ALSDE. Analyses and report writing were completed during the month of August.

### **Method**

The following data sources were used to develop the database:

- 2014-2015 annual reports from four active SGOs: Scholarships for Kids, AAA Scholarship Foundation, Alabama Opportunity Scholarship Fund, and Rocket City Scholarship Granting Organization.
- Supplemental demographic information provided by the SGOs that included gender, grade level and race.
- 2014-2015 test scores provided by the SGOs from participating schools. Test scores were received as PDFs and hard copies.
- 2014-2015 Alabama State ACT Aspire proficiency results available from the ALSDE website.

The four SGO annual reports list each student who received a scholarship and other information relevant to the student's award. These reports were combined to form a master list of scholarship recipients. The annual reports did not include key demographic information for each student (race, grade, and gender) and so the SGOs were asked to provide this information in a separate request. The test scores were initially entered into a separate data file by ISSR staff. For each scholarship recipient this included grade level, title of the test (e.g., ACT Aspire, Stanford Achievement Test), math scores, reading scores, and language arts scores (and/or a comparable score such as English). For each subject area, scale scores and national percentile scores were included if available. The SGO annual reports, demographic information, and test scores were then merged into a single file. This master data file was used to address evaluation questions.

### **Challenges**

Ideally, the current evaluation would synthesize the achievement test data for all scholarship recipients and conclusions would be drawn using a common metric across scholarship recipients and students attending Alabama public schools. However, the first evaluation of the scholarship program encountered several challenges related to reaching this ideal that hindered meeting the objectives of a) describing the learning gains of students in the scholarship program and b) making

comparisons between the learning gains of the scholarship recipients and comparable students attending public schools.

The first challenge concerned addressing learning *gains*, which would require a minimum of two years of test data. Test scores from 2013-2014 were available for only a small subset of 2014-2015 scholarship recipients (approximately 100 students). A request was made to the SGOs to collect 2013-2014 test scores from schools, and each SGO made a concerted effort to retrieve these scores. For a variety of reasons (e.g., records were missing, schools did not respond to requests due to summer recess), these data were not available for a large number of students. Reporting on the learning gains for only a small subset of students who had 2013-2014 scores was not appropriate because, due to the issues described below, the number of children in a single grade level did not meet the minimum number for reliable reporting. As a result, this report provides a description of the current level of academic achievement of scholarship recipients, rather than their gains over time. This information is still pertinent for assessing the impact of the AAA on student achievement, especially since achievement levels can be compared to national norms on most tests. Subsequent reports will attempt to describe learning gains as a more complete database of yearly test scores is compiled.

A second challenge is due to the use of 18 different standardized tests by the 104 reporting schools. The AAA allows for schools to use any nationally normed test. An essential requirement in translating scores across tests is that the tests must seek to measure identical traits, skills, and abilities, such that a student would not perform differently between the tests. However, tests vary in their content and are designed for unique purposes, which make comparisons across tests invalid. The ACT Aspire test, for example, is a criterion-referenced test, and scores describe student success in meeting achievement readiness benchmarks that indicate if the student is on track to meet college entrance requirements. In theory, 100% of the students in Alabama could achieve these criterion benchmarks. In contrast, tests such as the Stanford Achievement Test 10, the Iowa Test of Basic Skills, and Terra Nova are norm-referenced tests that are designed to compare student achievement relative to others at a particular grade level and distinguish between high and low achievers. In brief, criterion-referenced and norm-referenced tests can be quite different in their design. Even comparisons across norm-referenced tests are difficult to make because they may test different skills, or the same skills in different fashions. There are no national studies, of which we are aware, that have equated the 18 different standardized tests that were used by the schools receiving scholarship recipients. As a result, although all tests are valid instruments for assessing student academic achievement, scores cannot be combined or directly compared across the different tests. In this report student test results are reported separately for each test.

A third and related challenge concerns comparing the academic achievement of scholarship recipients to students attending Alabama public schools. Due to the aforementioned issues related to the comparability of different tests, only scholarship students who took ACT tests were statistically compared to the Alabama public school students (who also take ACT tests). Even then, caution must be taken in drawing conclusions based on these analyses because there is no way to determine if scholarship students with ACT scores are representative of the larger group of scholarship recipients. For example, schools that give the ACT Aspire might have different academic curricula (e.g., focusing on college entrance) compared to those that give the Stanford Achievement Test. Small numbers for some grade levels and demographic groups also make comparisons potentially unreliable. Guidance from ACT recommends a sample of at least 25 students, and this is a standard that was adopted in this report.

Finally, comparisons between scholarship recipients and public school students was limited to examining the percentage of students falling into the four proficiency groups related to the ACT readiness benchmarks (*in need of support*, *close*, *ready*, and *exceeding*) utilized in the Department of Education's data reporting. Each of these groups represents a range of scores and meaningful variation within each of the four groups could not be detected. For example the level 1 proficiency group for 8<sup>th</sup> grade math has scale scores that range between 400 and 418. If the average score of one group of students were 400 and another were 418, they would both fall in the same proficiency group, yet being close to the benchmark may be a noteworthy indication of progress. Conversely, if two groups had average scores of 418 and 419 respectively, they would be in different proficiency groups, and yet their test outcomes may not be significantly different. Replacing the proficiency groups with the full range of scores would be more sensitive to detecting differences between the two groups. An average score can be calculated for the scholarship recipients, but data were not available to create a comparable score for the students in the public schools. ALSDE was asked if State data were available in another form that would permit more precise comparisons (i.e., that would allow means to be calculated), but they were not able to provide data in any other format than proficiency groups.

As a result of these various challenges and limitations, this first report cannot accomplish all that was desired. Thus, the report will end with several recommendations for improving the quality of the data available for conducting future evaluations.

### **Demographic Description of Scholarship Recipients**

SGO annual report data were used to develop a description of students who were scholarship recipients. ISSR was able to identify 4115 students from these reports in kindergarten through 12<sup>th</sup> grade who received scholarships during the 2014-2015 academic year. Because in some cases the method by which the annual reports accounted for scholarship recipients was at odds with the reporting format needed for the evaluation, this number may underestimate the true number of scholarships, though not substantially so. The majority of the students were first time scholarship recipients (77%), free/reduced lunch eligible (98%), and were identified as low income eligible based on the AAA criterion (89%). The racial make-up of the scholarship recipients was 66% Black/African American, 18% White/Caucasian, 8% Hispanic, and 8% were another race or no information was provided. Gender was nearly evenly divided between males (49%) and females. Students resided in 38 unique counties in the state of Alabama and approximately 32% were zoned to attend a failing school. Students in grades kindergarten through second grade comprised 1195 (29%) of scholarship recipients and were not required to report test information.

### **Scholarship Recipients with Test Data**

One hundred and four (104) schools provided test scores for 1777 students. Due to the limited amount of time available for creating this first report, ISSR was not able to determine why the number of tests received fell short of the number expected based upon the SGO annual reports. Several factors resulted in further reducing the number of student test scores that were included in the evaluation of the outcomes.

- Two-hundred and fifty-six (256) students were in kindergarten through second grade, grades for which standardized testing was not required by ALSDE for 2014-2015.
- Of the tests that were given to students in grades 3 or higher, twelve schools (135 students) used unique tests that no other school used. These schools typically had a low number of scholarship recipients. Making public these test results (especially when aggregated by grade,

race, or gender) would potentially lead to undesirable results: a) It risks creating a situation in which schools and individual children could be potentially identifiable; the latter a violation of FERPA regulations; b) Small samples, as noted earlier, are not likely to be representative of the full group of scholarship recipients, and consequently, results from these samples would not contribute meaningfully to the evaluation of AAA.

- An additional 145 test scores provided by the schools and SGOs for students could not be matched to the names listed in the SGO annual reports. Considerable effort was made to identify the children in the SGO records. The inability to match these students could be due to schools inadvertently including test scores for students not receiving scholarships or to errors made in the test forms or SGO records (e.g., misspelled names, misreporting grade levels). Once these students were eliminated from the test score data base, an additional test, PSAT, no longer had a sufficient number of students (20) to contribute reliably to the evaluation.
- Finally, several schools provided copies or PDFs of test scores that were illegible (in part or whole), resulting in a further reduction in the sample. In some cases we were able to work with the SGOs to get new copies of the tests, but some schools could not be reached due to the summer break.

The remaining test score data included nearly a thousand students from 91 schools. Data provided from five standardized tests are included in this report: 1) ACT Aspire (also used by ALSDE); 2) The Stanford Achievement Test 10; 3) Terra Nova; 4) The Iowa Test of Basic Skills; and 5) The Practice ACT (college entrance exam). The Table below indicates the number of students associated with each test. Further attrition occurred for specific tests due to missing test scores. (Schools may not have included a particular subject area in their reports or individual students may not have tested in a subject area).

<b>Tests Included in the Evaluation</b>		
Test	Number of Students	Number of Schools
ACT Aspire	137	9
Stanford 10	431	52
Iowa Test	275	19
Terra Nova	52	4
Practice ACT	75	7
<b>Total</b>	<b>970</b>	<b>91</b>

Nearly all of the tests provided by the schools purport to base their test questions on nationally recognized educational standards, such as those of the National Assessment of Educational Progress (NAEP) and provide a score, such as a national percentile, that can be used to evaluate student performance relative to other students in the U.S. These tests could be used to assess whether students or school systems have met requirements set by national or state standards, and thus, meet the testing requirement put forward in AAA. A brief description of each of the five tests follows.

- The *ACT Aspire* assesses progress toward college and career readiness. Benchmarks provided by ACT are used for understanding if a student is on track to succeed in entering college. Scale scores indicate students' performance against a set of learning standards for each grade level.

As such, ACT Aspire scores are criterion-referenced, and it is possible for every child to get a score that meets the benchmark. ALSDE has adopted the benchmark scores used by ACT Aspire to create four proficiency levels: *In need of support* (Level 1), *Close* (Level 2), *Ready* (Level 3), and *Exceeding* (Level 4). Students scoring at Level 3 or higher are considered proficient. National percentile scores comparing students' scores relative to other students in the country at the same grade level are provided by ACT. Unlike the benchmark categories, the percentile scores are not readily interpretable as to whether a child is meeting learning standards for their grade. The ACT Aspire includes test scores for reading, language arts, English, and mathematics in addition to other areas.

- The *Stanford Achievement Test 10* is a norm-referenced test and was developed, among other reasons, to provide a way to compare a child's academic achievement relative to others in the nation. The scale scores follow a bell curve, or normal distribution, such that a child who scores at the 50<sup>th</sup> percentile is performing as well as or better than half of the students in the nation who are at the same grade level. In contrast to the ACT benchmarks, alone the percentile scores do not indicate if a child has acquired the academic skills and content that are appropriate for his or her age group. The Stanford provides achievement/ability scores in language, reading and math.
- *Terra Nova, 3<sup>rd</sup> edition* is a norm-referenced test similar to the Stanford Achievement Test. The test content aligns with the framework of the NAEP. The national percentile scores indicate how well a child compares to other students at the same grade level, similar to the Stanford Achievement Test. Included in the report are scores for language, reading, and math.
- *Iowa Test of Basic Skills* was developed by the Education Department at the University of Iowa and is a norm-referenced test. Test items were developed to align with the Iowa Core of State Educational Standards. The test has been validated at the national level, and it provides national percentile scores for reading, English and math.
- *The Practice ACT* is used to prepare high school students to take the College ACT. The scores can be used to predict how well a student might perform on the ACT college entrance exam. Reports include an estimated ACT score (1 - 36) and a national percentile score. Subscale scores are provided for reading, English and math.

### **Demographic Information for Scholarship Recipients Included in the Evaluation**

Based on information provided by the SGOs, the 970 scholarship recipients with usable test scores were largely first time scholarship recipients (94%) and eligible for free or reduced lunch (98%). Forty-three percent (43%) were zoned to attend a public school that was designated as failing by the ALSDE. The racial make-up of the sample was 65% Black/African American, 22% White/Caucasian, 5% Hispanic, and the remaining 8% of students were either another race or none was designated. The sample was evenly divided between male and female students, although for 8 students gender was not provided. Students represented 25 counties in the state and were in grades 3 - 12. The students with test score data were very similar to the larger sample on most demographic characteristic except that they had a greater percentage of first time scholarship recipients (94% vs. 75%).

### **Objective 1: Describe the Academic Achievement of Students in the Scholarship Program**

Outcomes for each of the five tests are described separately below. For each test, a brief description of the demographic information of the students is provided and additional test details relevant for understanding the test scores. When possible, test results aggregated by grade, race, and gender

are presented in tables. National percentile scores are included for every test. Presenting scale scores was sometimes useful in interpreting the test score information, and so these are included when relevant. The presentation of the results is organized by the type of test, norm or criterion-referenced. The first three tests, Stanford Achievement Test 10, Iowa Test of Basic Skills, and Terra Nova, are norm-referenced tests. Since these tests measure achievement in similar ways, findings are summarized for this group of tests together. The criterion-referenced Practice ACT and ACT Aspire are summarized next. The AAA asks for comparisons to be made for math and language arts subject areas. For some tests, English scores were provided rather than language arts, but the content of these subjects was similar. Furthermore, because the State of Alabama uses reading scores to evaluate public school students, reading scores are included as well in this report.

#### *Stanford Achievement Test 10*

The Stanford Achievement Test 10 was given to students in grades 3 through 12, but a sufficient number of test scores (25 or more) were available only for grades 4 - 8. Students who took the Stanford Achievement Test were predominantly first time scholarship recipients (98%) and free/reduced lunch eligible (98%). The racial make-up was 61% Black/African American, 28% White/Caucasian, 3% Hispanic, and the remaining 8% either had no information on race or were classified as another racial group. Gender was nearly evenly split between males and females. The table below provides the mean scale scores and national percentile scores for students in each grade. Some schools used 2002 norms in reporting the national percentile scores, despite the availability of more recent 2007 norms. The better indicator of student achievement relative to national standards is the 2007 norms, and consequently this standard was used in this report. It should be noted that the two tests are identical, and only the national percentile scores are different. The scale scores are included in the table as a reference point for the two sets of norms.

The only racial ethnic group with 25 or more students at any grade level was Black/African American, and only grades 4 through 8 met the minimum of 25 students. These are also presented in the table below. There was a sufficient number of male and female students to report test results for grades 4 and 6 through 8. For 5<sup>th</sup> grade there was a sufficient number of female, but not male students.

<b>Mean Stanford Achievement Test 10 Scores Grades 4 – 8 (2007 Norms)</b>							
Grade	Group (N)	Math		Reading		Language	
		Scale Score	Percentile	Scale Score	Percentile	Scale Score	Percentile
4	All (61-62)	623	43	631	37	629	42
	AA/Black (36-37)	616	36	621	29	621	35
	Female (30-31)	628	48	638	44	643	57
	Male (30)	619	39	625	32	616	31
5	All (51-52)	636	32	632	22	631	30
	AA/Black (31-32)	630	27	626	17	622	22
	Female (30)	635	31	632	22	636	34
	Male (< 25)	*	*	*	*	*	*
6	All (88-90)	649	32	647	25	641	31
	AA/Black (57-59)	644	28	639	19	634	24
	Female (43)	653	36	657	34	654	48
	Male (44-46)	646	30	637	18	629	20
7	All (65-66)	663	35	660	28	650	30
	AA/Black (39-41)	649	21	648	18	638	22
	Female (37-38)	660	32	660	28	653	33
	Male (26-28)	665	37	657	25	645	26
8	All (68-73)	666	30	672	32	655	30
	AA/Black (43-48)	660	24	669	29	652	28
	Female (40-43)	665	29	672	32	662	36
	Male (28-30)	669	33	671	31	646	23

\* Indicates an insufficient number of students in the group (< 25) for reporting.

*Iowa Test of Basic Skills*

The Iowa Test was administered to students in grades 3 through 8, but there were reports for only 18 students in 3<sup>rd</sup> grade, failing to meet the minimum requirement of 25 students. The table below provides the results for the remaining grades. Students taking this test were generally first time scholarship recipients (94%) and free/reduced lunch eligible (99%). Females comprised 54% of the test takers. The racial make-up was largely Black/African American (65%), White/Caucasian (17%), and Hispanic (11%). The only racial group with 25 or more students at any grade level was Black/African American, and only for grades 6 and 7. Similarly, there were sufficient numbers of female students to report test results for grades 6 and 8; however, for males only 6<sup>th</sup> grade had a sufficient number of students.

<b>Mean Iowa Test Scores Grades 4 – 8</b>				
Grade	Group (N)	Math Percentile	Reading Percentile	English Percentile
4	All (34-39)	29	41	36
	Black (< 25)	*	*	*
	Female (< 25)	*	*	*
	Male (< 25)	*	*	*
5	All (26-29)	41	43	47
	Black (< 25)	*	*	*
	Female (< 25)	*	*	*
	Male (< 25)	*	*	*
6	All (68-79)	33	41	41
	Black (52-55)	28	36	37
	Female (27-28)	26	34	38
	Male (25-28)	30	37	36
7	All (64-81)	35	44	43
	Black (48-59)	31	38	38
	Female (< 25)	*	*	*
	Male (< 25)	*	*	*
8	All (25-26)	30	39	*
	Black (< 25)	*	*	*
	Female (26-36)	29	42	44
	Male (< 25)	*	*	*
* Indicates an insufficient number of students in the group (< 25) for reporting.				

*Terra Nova*

Terra Nova test score results were available for 52 students in grades 3 through 12. None of the grade levels had an adequate number of students to provide reliable test results. The table below provides the results for all grades combined. Sixty-four percent (64%) of the students were first time scholarship recipients and 94% were free/reduced lunch eligible. The group was composed of 69% males and was predominantly Black/African American (94%), with the remaining 6% classified as White/Caucasian. The only racial group with a sufficient number of students was

Black/African American. It should be noted that because only a few test takers were not in this racial group, these scores are nearly identical to the total scores for all students. A breakdown of scores by gender is also reported in the table below; however there were not enough female participants to provide reliable scores.

Mean Terra Nova Test Scores Grades 3 - 12 Combined				
Grades	Group (N)	Math Percentile	Reading Percentile	Language Percentile
3-12	All (52)	45	44	42
	Black (49)	44	41	39
	Female (< 25)	*	*	*
	Male (36)	48	48	43
* Indicates an insufficient number of students in the group (< 25) for reporting.				

### Summary for Stanford 10, Iowa Test of Basic Skills, and Terra Nova

In interpreting norm-referenced tests, it is important to be mindful that the percentile scores are an assessment of students' performance relative to other children at the same grade level in the country. By themselves, the scores do not indicate if a child has acquired the knowledge and skills expected for their grade. Consequently, although the 50<sup>th</sup> percentile is often used as the yardstick for evaluating performance, it is not a good indicator of a child or a group of children having mastered grade level material. As a marker for performance, however, the scholarship recipient mean scores should be close to the 50<sup>th</sup> percentile, if *as a group* they are achieving at levels similar to others in the U.S. Generally, meeting or exceeding this standard could be considered an acceptable outcome. Separately for each test, statistical comparisons were made to assess if the mean score for the scholarship recipients in each subject area was significantly higher or lower than the 50<sup>th</sup> percentile. Considering first the Stanford 10, the average scores for each grade level and subject area were generally significantly below the 50<sup>th</sup> percentile. The exceptions were for 4<sup>th</sup> grade math and language, which were not significantly different than the 50<sup>th</sup> percentile score. For the Iowa Test, the mean national percentile scores were significantly below the 50<sup>th</sup> percentile for all grades except 5<sup>th</sup> grade. In contrast to the Iowa Test and the Stanford 10, none of the mean percentile scores for Terra Nova were significantly below the 50<sup>th</sup> percentile, suggesting a better outcome for these students.

It is not clear why the results for Terra Nova were different compared to the other tests. Discrepancies could be due to differences in the content of the tests, differences in the schools that happened to choose one test over another that may produce higher levels of achievement (e.g., curriculum, pedagogical approaches), or due to different student demographic make-up. Compared to the other tests, students who took the Terra Nova had a greater proportion of students who were *not* first time scholarship recipients (37%), were Black/African American, (94%), and were male (69%). The association between the latter two factors and higher test scores has generally not been demonstrated. As this is the first year of the evaluation, no conclusions can be drawn about whether the AAA scholarship program is associated with higher learning gains the longer a student participates. Given that the Stanford 10 and Iowa Test include scores for nearly 600 students and Terra Nova includes only 52 students, the preponderance of the data indicates that the scholarship

recipients were generally performing lower than the typical student in the U.S. This finding is similar to that of the NAEP analysis of a representative sample of students in the state of Alabama indicating that Alabama students lag behind other states in achievement.

*ACT Practice Test*

The ACT Practice Test was administered to 75 students in grades 9 through 12. Nearly all students were first time scholarship recipients (97%) and were eligible for free/reduced lunch (99%). Black/African American students comprised 80% of the test takers, with White/Caucasian students representing the next largest group (12%). The remaining 8% of students were either from another racial group or had no information. Males comprised 57% of the students. Only grade 11 met the minimum requirement of 25 students for reporting. Both percentile scores and scale scores are presented because scale scores are comparable to the ACT College Entrance Exam scores (range 1 - 36). The only racial ethnic group with 25 or more students in grade 11 was Black/African American (34 students, 85%). The remaining students were classified as White/Caucasian. There were not enough students to breakout the scores by gender.

Mean ACT Practice Test for Grade 11							
Grade	Group (N)	Math		Reading		English	
		Scale Score	Percentile	Scale Score	Percentile	Scale Score	Percentile
11	All (36-40)	15	16	15	20	13	19
	AA/Black (30-33)	15	14	14	19	13	17

Note. There was an insufficient number of students to breakdown the scores by gender.

Across the board, the average ACT scores fell below benchmark scores for college preparedness set by the State of Alabama (18 for English, 22 for Math, and 22 for Reading). In this case, the benchmark score is the minimum score needed on an ACT subject-area test to indicate a 50% chance of obtaining a B or higher or about a 75% chance of obtaining a C or higher in the corresponding college course. Proficiency groups were created based on the benchmark scores. For reading and math, only 5% of the students met or exceeded the benchmark scores, and for English, 15% met or exceeded the benchmark score. Together these results suggest that the scholarship recipients who took the ACT Practice Test failed to meet national standards predictive of college achievement.

*ACT Aspire*

The ACT Aspire test was administered to 137 students in grades 3 through 10. Similar to the other test groups, most students were first time scholarship recipients (93%) and were eligible for free/reduced lunch subsidies (96%). Students were 64% Black/African American, 23% White/Caucasian, 4% Hispanic, and the remaining 9% were in another ethnic group or not designated. Females comprised 53% of this group. Only grades 6, 7, and 10 met the minimum reporting requirement of 25 students. There were insufficient numbers of students in these grade levels to present scores aggregated by race or gender.

As described earlier, the ACT Aspire scoring includes four proficiency benchmarks that classify students as 1) *in need of support*, 2) *close*, 3) *ready*, and 4) *exceeding*. The proficiency benchmarks for reading and math are the same benchmarks that the ALSDE uses to assess student achievement

in public schools. Students who are at and above the benchmarks are on target for college readiness by the time they are in the 11th grade. Although an effort was made to report language arts or English achievement as indicated in the AAA, these scores were not reliably provided by the schools for grades 3 through 8, and there were insufficient numbers of students with these scores at most grade levels. (Grade 10 English results are presented in the section comparing scholarship recipients to students in Alabama public school students.) As a result, the tables below include math and reading scores for grades 6, 7, and 10. Additionally the mean percentile scores and the percentage of students in each of the proficiency groups were calculated for the sample as a whole (4<sup>th</sup> - 10<sup>th</sup> grades) as a way to represent the performance of more of the students and to provide further insight into the achievement of scholarship recipients. (Third graders were not included because scale scores, which are used to calculate proficiency scores, were not provided for these students.)

The average national percentile ranking is provided in the first table that follows, and the second table presents the percentage of students falling into each of the four proficiency groups (based on 2015 norms). For the combined grades 4 - 10 there were sufficient numbers of Black/African American and White/Caucasian students to break out the scores by these two racial groups. In addition, scores are reported for male and female students.

<b>Mean ACT Aspire Scale and National Percentile Scores (2015 Norms)</b>						
<b>Math</b>				<b>Reading</b>		
Grade (N)	Math Scale Score	Corresponding Proficiency Level	Math Percentile	Reading Scale Score	Corresponding Proficiency Level	Reading Percentile
6 <sup>th</sup> (30-31)	418	2	47	418	2	45
7 <sup>th</sup> (25)	418	2	47	420	2	57
10 <sup>th</sup> (26-27)	425	1	48	422	2	51
4 <sup>th</sup> -10 <sup>th</sup> (132-133)	----	----	48	----	----	53

Proficiency Levels: 1 = *In need of support*, 2 = *Close*, 3 = *Ready*, 4 = *Exceeding*

<b>ACT Aspire Mathematics and Reading: Number and Percentage of Students in Proficiency Groups</b>									
Grade (N)	Math Proficiency Group				Reading Proficiency Group				
		1 Needs Support	2 Close	3 Ready	4 Exceeds	1 Needs Support	2 Close	3 Ready	4 Exceeds
6 <sup>th</sup> All (30-31)	%	13%	52%	26%	10%	33%	47%	17%	3%
	n	4	16	8	3	10	14	5	1
7 <sup>th</sup> All (25)	%	36%	40%	24%	0%	28%	32%	32%	8%
	n	9	10	6	0	7	8	8	2
10 <sup>th</sup> All (26-27)	%	54%	19%	23%	4%	48%	15%	37%	0%
	n	14	5	6	13	13	4	10	0
<b>4th-10<sup>th</sup></b>									
All (123-124) <sup>1</sup>	%	29%	42%	25%	4%	32%	32%	29%	7%
	n	36	52	31	5	39	39	36	9
Black/AA (78)	%	31%	49%	18%	3%	38%	36%	24%	3%
	n	24	38	14	2	29	27	18	2
White/CA (26)	%	23%	35%	35%	8%	19%	30%	41%	11%
	n	6	9	9	2	5	8	11	3
Female (65-67)	%	24%	46%	25%	5%	34%	29%	28%	9%
	n	16	31	17	2	22	19	18	6
Male (56-57)	%	34%	38%	25%	4%	28%	35%	32%	5%
	n	19	21	14	2	16	20	18	3

<sup>1</sup> None of the 3<sup>rd</sup> grade students had scale scores and consequently they could not be classified into a proficiency group. Four students were missing math scale scores and could not be classified into a proficiency group. Five students were missing reading scale scores and could not be classified into a proficiency group.

*Note.* Groups 3 and 4 have met or exceeded grade level benchmarks. Only grades 6, 7, and 10 met the minimum sample size (25) to be reported.

The ACT Aspire is a criterion-referenced test, and as a result, the national percentile scores are interpreted differently compared to norm-referenced test. Across the different age groups, the national percentile scores for math and reading are between 45 and 57. Although these rankings might appear higher than those previously reported for the norm-referenced tests, they correspond to scale scores that indicate a non-proficient level of achievement. The majority of scholarship students taking the ACT Aspire at each grade level fall below the benchmarks for their grade level for math and reading. When proficiency scores are collapsed across grades 4 through 10, only the White demographic group has more than 50% of the students scoring in the proficient range, and only for reading.

## **Summary of the ACT Practice Test and ACT Aspire Test**

Across both of these criterion-referenced tests, scholarship recipients' mean scores fell below proficiency levels. Further clarity on the significance of these results is found by comparing these scores to other students attending public schools in Objective 2.

### **Objective 1 Conclusions**

Without knowing the previous year's achievement scores, it is impossible to state whether the scholarship recipients' academic achievement represents an improvement, decline, or no change as a result of the AAA. Collectively, across the different tests and grades the scholarship recipients fell below national norms in all subject areas. This results needs to be placed in the context of the academic achievement of students in the state as a whole, as many indicators, such as the NAEP assessments and the ALSDE annual reports, indicate that students in the state of Alabama fall short of national norms and benchmarks. Additionally, the students in the AAA program belong to demographic groups (low income, racial minority groups) that have lagged behind other students in the state and in the country in academic achievement. As a result, the first evaluation of this objective cannot draw a definitive conclusion about the impact of the program on the academic performance of the scholarship recipients. A later section in this report will identify steps to be taken in the future to improve the evaluation of the AAA.

### **Objective 2: Compare Scholarship Recipients to Alabama Public School Students**

To meet Objective 2, comparisons were made between scholarship recipients and the scores reported by the ALSDE for the ACT Aspire (3<sup>rd</sup> - 8<sup>th</sup> grades), ACT Plan (10<sup>th</sup> grade) and the ACT college entrance exam (11<sup>th</sup> grade). Alabama uses reading and math proficiency scores in grades 3 - 8 and English and math proficiency scores for grade 10 to assess the academic achievement of students in public schools. It should be noted that the 10<sup>th</sup> grade students in public schools were administered the ACT Plan in Spring 2015, rather than the ACT Aspire. Information from the ACT Aspire support website indicates that valid comparisons between the two tests can be made by comparing the percentage of students achieving the benchmarks for each test. Nevertheless, it is important to be mindful that the two tests are not identical. In the coming years, ALSDE will use the ACT Aspire for 10<sup>th</sup> graders, eliminating this issue in the future. The Practice ACT (used by scholarship recipients) and the ACT college entrance exam are scored on the same scale.

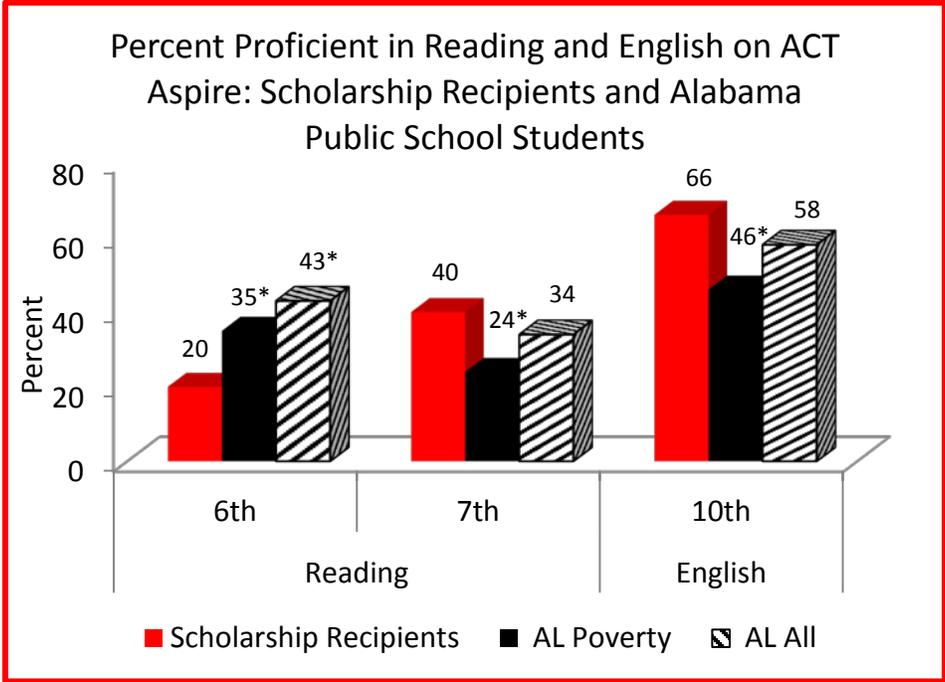
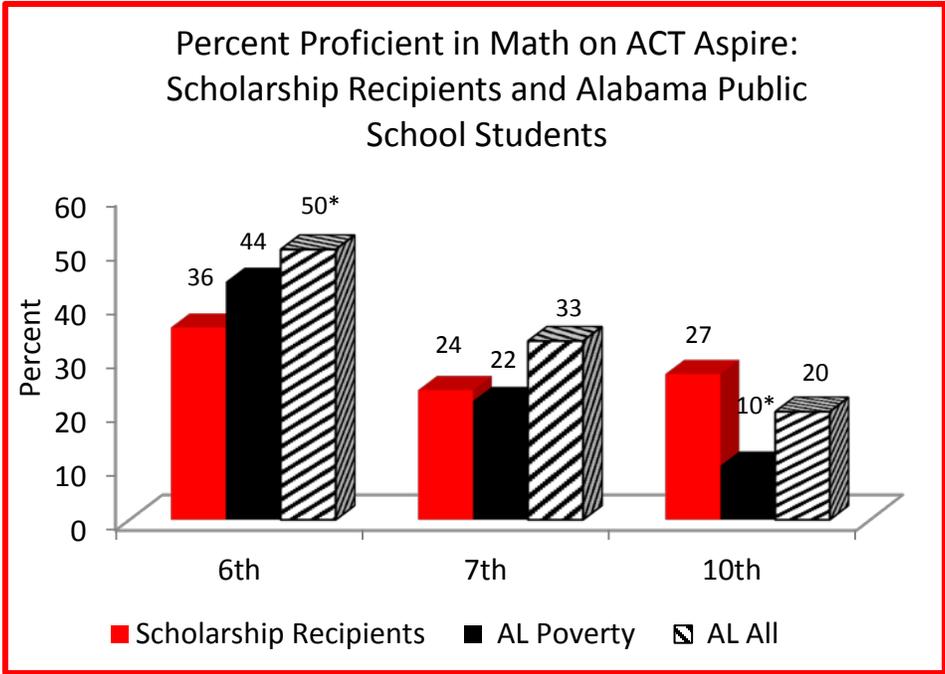
As previously noted, the comparison between scholarship recipients and students attending public schools was limited to the proficiency scores due to the limited availability of ALSDE data. The reliability of the comparisons was further hampered by the small number of scholarship students with usable ACT Aspire scores (133, grades 3 - 10) and Practice ACT scores (40, grade 11). A valid concern is that this small subsample of students is not representative of the participants in the AAA program. There may be factors associated with schools that give the ACT Aspire test (as opposed to other tests such as the Stanford 10) that make them unrepresentative of the schools with scholarship recipients (e.g., larger class sizes, teacher qualifications, pedagogical approaches).

With these limitations in mind, the available data were used to make the comparisons that are set forth in the evaluation requirements for the AAA. The charts below indicate the distribution of scholarship and public school students across the four proficiency groups for grades 6, 7, and 10. ALSDE statistics are provided for all students at each grade level and for those classified as "poverty" based on being eligible for free/reduced lunch subsidies. The poverty group provides

the most appropriate comparison group, as nearly all scholarship recipients are eligible for the subsidy. The percentage of children achieving proficiency (at levels 3 and 4) is represented graphically as well. The State evaluates 10<sup>th</sup> grade students based on their English scores, not reading, and this difference is reflected in the table below.

<b>Mathematics and Reading: Percentage of Scholarship and Alabama Public School Students in Proficiency Groups</b>								
Grade/ Group	<b>Math Proficiency Group</b>				<b>Reading Proficiency Group</b>			
	1 Needs Support	2 Close	3 Ready	4 Exceeds	1 Needs Support	2 Close	3 Ready	4 Exceeds
<b>6<sup>th</sup> Grade</b>								
Scholarship	13%	52%	26%	10%	33%	47%	17%	3%
AL Poverty	14%	42%	31%	13%	30%	35%	22%	13%
AL All	14%	36%	32%	18%	29%	28%	25%	18%
<b>7<sup>th</sup> Grade</b>								
Scholarship	36%	40%	24%	0%	28%	32%	32%	8%
AL Poverty	39%	40%	17%	5%	42%	34%	20%	4%
AL All	30%	37%	22%	11%	33%	33%	27%	7%
	<b>Math Proficiency Group</b>				<b>English Proficiency Group</b>			
	1 Needs Support	2 Close	3 Ready	4 Exceeds	1 Needs Support	2 Close	3 Ready	4 Exceeds
<b>10<sup>th</sup> Grade</b>								
Scholarship	54%	19%	23%	4%	21%	13%	33%	33%
AL Poverty	45%	45%	6%	4%	25%	30%	27%	19%
AL All	34%	46%	11%	9%	18%	24%	28%	30%

Statistical analyses compared the proportion of scholarship recipients who achieved or exceeded proficiency (levels 3 and 4 combined) at each grade level and subject area to the comparable scores for those who attended public schools. In the graph below, an asterisk (\*) above the bar for *Alabama Poverty* or *Alabama All* indicates that there was a significant difference between the percentage of students who were proficient in that group compared to the scholarship recipients. The comparison between the Alabama Poverty group and the scholarship recipients provided a mixed pattern of results. In two cases, math achievement in 6<sup>th</sup> and 7<sup>th</sup> grade, there were no significant differences. In three cases, the percentage of proficient scholarship recipients was higher than the state poverty students: 7<sup>th</sup> grade reading, 10<sup>th</sup> grade math, and 10<sup>th</sup> grade English. In one case, 6<sup>th</sup> grade reading, the scholarship recipients were less proficient. The comparison between the Alabama All group and the scholarship recipients was similarly mixed. Comparisons for 7<sup>th</sup> and 10<sup>th</sup> grade students yielded no significant differences; where as both comparisons for 6<sup>th</sup> grade students indicated that there were fewer scholarship recipients in the proficient category.



\* Indicates that the mean score for the group is significantly different from the mean scholarship recipient score.

ACT college entrance exam scores were compared for 11<sup>th</sup> grade scholarship students who took the Practice ACT and the ALSDE report of 11<sup>th</sup> grade ACT scores. As noted earlier, only 5% of the scholarship students met or exceeded the benchmark scores for math and reading, and 15% met or exceeded the benchmark score for English. For 2015, ALSDE reported that state-wide the

percentage of students meeting the benchmarks was 19% for reading, 28% for math, and 45% for English. State scores for students receiving free/reduced lunch could not be located. Based upon the data available, the scholarship recipients appear to be less proficient compared to other 11<sup>th</sup> grade students in the public schools.

## **Objective 2 Conclusions**

The strength of any conclusions for Objective 2 relies on how representative the students included in the analyses are of the larger group of scholarship recipients. Information is lacking with respect to a number of grade levels, and the number of students overall is relatively small. Additional concerns have been presented with respect to unknown differences among schools that choose particular tests. Unfortunately, it is impossible to address these limitations and concerns adequately with the information available. As a result, the summary presented here is based on the best available information, but is inconclusive. On the other hand, it lays the groundwork for future evaluation of the program.

The comparisons at each of the grade levels with adequate test information (6<sup>th</sup>, 7<sup>th</sup>, 10<sup>th</sup>, and 11<sup>th</sup>) did not present a cohesive pattern. Because the scholarship recipients come from low income families and are generally free/reduced lunch eligible, the most appropriate comparison demographic group from the state is the “poverty” group. Sixth grade scholarship recipients performed similarly to their public school counterparts in math, but worse in reading. Seventh grade scholarship recipients also performed similarly to public school students in math, but *better* in reading. Tenth grade high school students performed better than public school students in both math and English; yet 11<sup>th</sup> grade students performed worse on the ACT college entrance exam. For both groups of students, there are very few subjects in which proficiency levels reached 50% or higher in 2015, suggesting that there is need for improvement across all schools. *There is no compelling evidence to suggest that scholarship recipients as a group performed differently than their counterparts attending public schools, and the inability to provide compelling evidence is largely due to limitations in the information that was available.* The ability to draw stronger conclusions in future evaluations will depend on the number of schools with scholarship recipients that use the ACT Aspire and other tests that are mandated by ALSDE for public schools. This will allow for confident comparisons across the same test, with a larger and more representative sample.

## **General Conclusions**

At the heart of the evaluation is the question of how the scholarship program enacted through the AAA affected the academic achievement of students in the program. Throughout this report a range of factors have been cited that limited the ability of the evaluation to adequately answer this question. Nevertheless an attempt was made to make the best use of the information available to describe how well the scholarship recipients performed. By most metrics, the 2015 scholarship recipients as a group fell below national norms and benchmarks on standardized tests, as do their counterparts in the public school system. The vast majority of scholarship recipients had only been at their alternative school for a year, that is, they were first time scholarship recipients. The impact of the alternative schools will be clearer as multiple years of test scores become available.

## **Improving the Evaluation**

Many challenges were encountered in conducting the evaluation of the AAA scholarship program, some of which have been noted throughout this report. ISSR will be working with the Department of Revenue and the SGOs to improve test score reporting and the collection of relevant

demographic information to increase the number of students included in the evaluation. In subsequent evaluations, test score data should be available that will allow for the analysis of achievement gains, as stated in the Act. In addition, three chief recommendations are made that would allow the evaluation team to better meet the objectives of the evaluation:

1. *Require all schools to administer to the scholarship students the same tests as the state administers to public school students.* This would provide a common metric across schools, yielding sufficient numbers of students to evaluate each grade and to make comparisons to the state reported outcomes. Recently, the publisher of the Stanford Achievement Test indicated that they would be phasing out the test. This may be an opportunity for many schools to switch to the ACT Aspire.
2. *Require new scholarship recipients to provide achievement test reports from prior years.* Students in grades 4 - 12 who previously attended a public school should have prior achievement test results in their academic records. Having these documents will allow the evaluation to consider academic gains for a larger group of students.
3. *Acquire Alabama State test score data that would allow for the calculation of mean scores, in addition to proficiency groups.* As noted earlier in the report, a great deal of information is lost with the use of proficiency scores. Proficiency scores essentially reduce actual scores to 4-point and 2-point scales, resulting in a loss of information. A better understanding of student academic gains can be achieved by using the full range of original test scores. Additionally, state test score data can only be displayed by a single demographic at a time. For instance, the proficiency scores of males can be examined, but the proficiency scores of white-males or black-males cannot. Further breakdowns, such as white-males who receive free or reduced lunch are also not possible. This lack of specificity can mask the disproportionate impact on specific groups that may benefit more or less from the program.

## Glossary of Terms

*ACT Readiness Benchmark Scores.* Achieving a score that meets or exceeds the benchmark scores indicates that a student has a high probability of future success in first-year college courses. There are four readiness groups (*in need of support, close, ready, and exceeding*) that correspond to a range of scale scores that are unique to each grade level.

*Criterion-referenced test.* These tests assess students' learning against a fixed set of predetermined learning standards that are set for their grade level. In an ideal school, every student would meet the criterion score for their grade level.

*Mean.* A mean test score is calculated by adding together every test score in a group and dividing by the number of people in the group. It is one way to represent the score of a typical person in the group.

*National percentile.* National percentile scores can range from 1 - 99. The percentile rank indicates the percent of students nationwide who scored lower than a particular raw score on the same test at the time the norms were compiled.

*Norm-referenced test.* These tests are designed to compare student achievement relative to others at a particular grade level with the goal of distinguishing between high and low achievers. National percentile scores are commonly used as a reference point for these tests, with the 50<sup>th</sup> percentile indicating the score achieved by the average student in the U.S.

*Proficiency Scores/Groups.* For the state of Alabama, proficiency scores correspond to the ACT readiness benchmarks defined above.

*Raw score.* A raw score is the number of items that a child answered correctly on a test.

*Scale(d) score.* A scaled score is a mathematical transformation of a raw score. Scaling provides a continuous metric across the different forms and levels of a test (such as tests for different grade levels). Higher scale scores indicate higher levels of academic achievement.

*Scholarship Granting Organization (SGO).* An organization that provides educational scholarships to eligible students attending qualifying schools. SGOs receive donations from individuals and corporations (subject to limitations imposed by the Alabama Accountability Act), which are then distributed in the form of scholarships to eligible students. Donations by taxpayers cannot be restricted or conditional with respect to how the donation is applied to scholarship recipients or schools.

*Statistically significant difference.* The difference between two scores is considered significantly different when there is a low probability (usually less than a 5% chance) that the difference could occur by chance. When a statistically significant difference is observed between the mean scores of two groups of students, it suggests that the difference is likely to be a "real" difference.

## Links to Online Resources

ACT Aspire Benchmarks

<https://www.discoveractaspire.org/assessments/standards-benchmarks/>

Alabama State Department of Education: Test outcomes for 2014-2015

<http://www.alsde.edu/dept/data/Pages/assessment-all.aspx>

Alabama State Education Report Card for 2014-2015

<http://www.alsde.edu/sec/comm/Pages/educationreportcard-all.aspx>

Alabama State testing requirements for 2014-2015

<https://www.alsde.edu/sec/sa/Testing/Student%20Assessment%20Timeline%20Revised%20September%202014.pdf>

Iowa Test of Basic Skills: General information

<https://itp.education.uiowa.edu/ia/default.aspx>

National Assessment of Educational Progress: Reports on student achievement

<http://www.nationsreportcard.gov/>

Norm- vs. criterion-referenced tests explained

<http://edglossary.org/norm-referenced-test/>

Stanford Achievement Test 10: General information

<http://www.pearsonassessments.com/learningassessments/products/100000415/stanford-achievement-test-series-tenth-edition.html#tab-details>

Terra Nova: General information

<http://www.ctb.com/ctb.com/control/ctbProductViewAction?productFamilyId=449&productId=733&p=products>