Evaluation of the Alabama Accountability Act: Academic Achievement Test Outcomes of Scholarship Recipients 2016 - 2017

The Institute for Social Science Research
The University of Alabama

Joan M. Barth, Ph.D.
Erika Steele, Ph.D.
Garrett Quenneville, M.A.

September 1, 2018

Executive Summary

This report fulfills the evaluation requirements of the 2013 Alabama Accountability Act by reporting on the academic achievement of the 2016-2017 scholarship recipients.

The report focuses on three objectives:

1. Describe the academic achievement of students in the scholarship program.
2. Compare scholarship recipients to Alabama public school students.
3. Assess changes in achievement across time.

Scholarship Granting Organizations provided demographic information and achievement test scores for scholarship recipients. 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:

- 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.
  - Some achievement tests were used by only one school or included only a small number of students, making analyses unreliable.
- The test score information available from the Alabama State Department of Education only includes the percentage of students in proficiency groups based on ACT Aspire and ACT College Entrance Exam scores, which limited the types of analyses that could be conducted.
- Inconsistencies in test score reporting from schools and missing test data limited the number of students who could be included in the evaluation sample.

The evaluation was based upon test scores from 1,991 scholarship recipients attending 114 schools in 43 counties. This represented 76% of the scholarship recipients in the grades for which testing was required. These students varied in their demographic characteristics:

- Number of years receiving a scholarship:
  - 15% were first time scholarship recipients.
  - 11% were two-time scholarship recipients.
  - 51% were three-time recipients.
  - 22% were in their fourth year.
- 90% were eligible for free/reduced lunch subsidies.
- 34% were zoned to attend a failing school.
- 62% were Black/African American, 20% were White/Caucasian, and 11% were Hispanic.

Continues

Although this report can show trends for this subsample of scholarship recipients, due to the necessity of excluding a significant proportion of scholarship recipients (24%) from analyses, findings may not be representative of all of the scholarship recipients.

Findings for Objective 1: Describe the academic achievement 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.
- Outcomes were even poorer for African-American participants who made up the majority of scholarship recipients (65%).
- These findings are similar to those of the National Assessment of Educational Progress for students attending public schools in Alabama.

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

- There were very few subject areas in which more than 50% of the students met proficiency standards for either group of students.
- For the ACT Aspire, comparisons did not present a clear pattern across subjects and grade levels to indicate that one group performed better or worse than the other.
- Overall, scholarship recipients in the 11^{th} grade performed about the same as their public school counterparts on the ACT.

Findings for Objective 3: Assess changes in achievement across time.

- On average, over time, participating in the scholarship program was not associated with significant improvement on standardized tests scores.
- On the ACT Aspire, students were more likely to remain in a non-proficient category than to improve. Although proficiency rates for 2016-2017 were higher for scholarship students than those of Alabama poverty students, the majority of students in both groups did not meet proficiency benchmarks.
- The overall lack of change over time follows the same pattern seen in public school students in Alabama and is likely not attributable to participation in the scholarship program.

Issues for Future Evaluations:

- Drawing conclusions regarding the academic achievement of scholarship recipients relative to students attending public schools depends on the number of schools with scholarship recipients that use tests that are utilized by ALSDE in the future.
- ALSDE discontinued the use of the ACT Aspire for the 2017-2018 academic year, and may change achievement tests again in 2018-2019. This will further constrain comparisons between scholarship recipients and students attending Alabama public schools over time.

Table of Contents

List of Charts, Tables, and Figures

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	Portable Document Format
PSAT/NMSQT	The Preliminary SAT/National Merit Scholarship Qualifying Test
r	Correlation coefficient
SGO	Scholarship Granting Organization

Evaluation of the Alabama Accountability Act: Academic Achievement Test Outcomes of Scholarship Recipients 2016 - 2017

Introduction

In September, 2016, the Institute for Social Science Research (ISSR) at the University of Alabama completed the first state-mandated evaluation of the academic outcomes of students receiving scholarships under the Alabama Accountability Act (AAA) as set forth in the AAA legislation. Thus far, in two previous reports, the ISSR has described the achievement test results for the 2014-2015 and 2015-2016 academic years and compared the outcomes to students attending public schools in Alabama. The current report follows a similar approach with the 2016-2017 achievement test results and additionally examines changes in scholarship recipients' achievement test scores over time in comparison to comparable children attending public schools in Alabama.

This report first provides an overview of pertinent AAA legislation. Next, the methodology is described including improvements from the 2017 report, continued challenges in data collection, and issues with compliance with the AAA. A summary of data collection efforts and findings for students receiving tuition scholarships in the 2016-2017 academic year is then provided, followed by comparisons to public school children, and an analysis of changes in test scores over time.

Overview of AAA

This report fulfills the evaluation component of the 2013 Alabama Accountability Act by providing evidence for the academic achievement of scholarship recipients in the 2016-2017 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 the standards set by the AAA. The AAA places restrictions on who can receive scholarships based on family income and school zoning. All students receiving scholarships must meet family income eligibility requirements. Priority is given to students who are zoned to attend a failing public school as designated by Alabama State Department of Education (ALSDE). However, students meeting AAA income requirements who attend non-failing public schools may receive scholarships if additional funds are available. Scholarships are awarded from the SGO to the student to attend a school that must meet standards set forth in the AAA. Scholarships may cover all or part of tuition and mandatory fees for one academic year. In 2015, the legislature amended the AAA 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 AAA.

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 include the following:

- The learning achievements of students receiving educational scholarships 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.
- 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.
- A report to be made every two years, starting in 2016.

Thus, the current 2018 report has three major objectives: a) describe the academic achievement of students in the scholarship program for the 2016-2017 school year, b) make comparisons between the learning achievement of the scholarship recipients and comparable students attending public schools for the 2016-2017 school year, and c) measure the learning gains of students in the scholarship program over time.

Alabama Mandated State Testing in Public Schools 2016-2017 Academic Year

Students attending public schools in Alabama during the 2016-2017 academic year were tested in March and April. Math and reading were assessed with the ACT Aspire for students in grades 3 - 8 and 10. Eleventh graders were required to take the ACT college entrance exam.

Method

Before the findings are presented it is important to recognize several challenges to meeting the evaluation objectives set forth in the AAA. As noted in the previous reports, the lack of a uniform achievement test among schools limits the conclusions that can be made about student learning gains. Scores from a total of 21 unique tests were provided. Comparisons across tests are invalid because tests vary in their content and are designed for unique purposes. Norm-referenced tests, such as the Iowa Test of Basic Skills and the Stanford Achievement Test, and criterion-referenced tests, such as the ACT Aspire, are based on different standards and cannot be directly compared.

Criterion-referenced test scores typically describe student success in terms of meeting achievement readiness benchmarks that indicate if the student is on track to meeting a long term academic goal, such as entrance to college. In theory, 100% of the students in Alabama could achieve these criterion benchmarks. In contrast, norm-referenced tests are designed to compare student achievement relative to others at a particular grade level and distinguish between high and low achievers. For example, a student scoring at the 70th percentile on a norm-referenced test achieved a score that was better than or equal to 70 percent of students in the nation at his or her grade level taking the same test. In criterion-referenced tests, the emphasis is on achieving scores that meet benchmarks, and consequently, percentile scores are less meaningful with respect to achievement. Even tests within the same broad categories of norm- or criterion-referenced cannot be combined for analyses since each test has unique content and unique scoring systems.

A second challenge was that some tests were used by only one school or included only a small number of students, making comparisons unreliable. Guidance from ACT recommends a sample of at least 25 students to achieve a reliable representation of students, and this standard was adopted in this report.

Inconsistencies in test reports provided by participating schools and missing data persisted as problems in the 2016-2017 school year. Schools did not reliably provide national percentile and scale scores for math, reading, and language arts/English, and missing data were often associated with specific test forms and schools. These inconsistencies in test score reporting and missing data compromise the integrity of the report findings. ISSR continues to work with the SGOs to communicate to the schools the information that is needed in the test reports.

With these challenges noted, the remainder of the report describes outcomes for the 2016-2017 academic year. Statistical comparisons were conducted throughout the report to aid in drawing conclusions. T-tests were used to compare the average scholarship student test scores to established benchmarks, to compare genders, or to compare racial/ethnic groups of scholarship students. Chi-square analyses were used to compare the proportion of scholarship students across different demographic (e.g., race) and proficiency groups. Z-tests were used to compare the percentages of scholarship students meeting benchmarks to comparable indicators of public school students. Correlations were used to assess the relation between achievement test scores and the number of years of participation in AAA scholarship program. These statistical tests take into account the sample size and the variation in the data to inform us of the likelihood of a reliable difference. As is customary in educational research, a probability value (p) of ≤ 0.05 was used as the criterion to determine significance.

Data Sources

The following data sources were used to evaluate the academic achievement of the 2016-2017 scholarship recipients:

- 2016-2017 demographic reports from the four active SGOs: Scholarships for Kids, AAA Scholarship Foundation, Alabama Opportunity Scholarship Fund, and Rocket City Scholarship Granting Organization.

- 2016-2017 test scores provided by the participating schools. The SGOs collected the majority of the test reports and then shared them with ISSR. Test scores were received as PDFs and hard copies.
- 2016-2017 Alabama State ACT Aspire proficiency results and the 11th grade ACT results available from the ALSDE website.
- 2014-2015 and 2015-2016 Alabama State ACT Aspire proficiency results.
- 2014-2015 and 2015-2016 achievement test scores from scholarship recipients.

2016-2017 Evaluation Sample

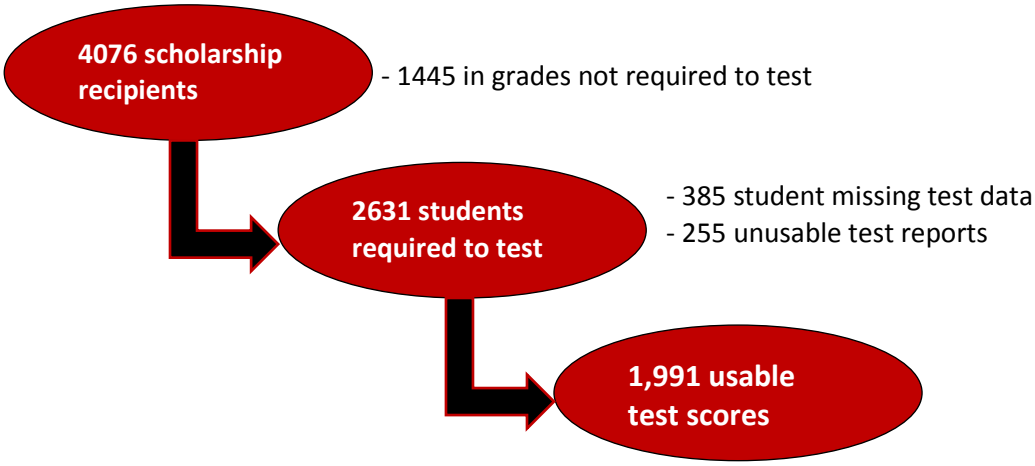
The SGOs reported that a total of 4076 students (50% female) in kindergarten through 12th grade had received scholarships during the 2016-2017 academic year. The majority of the students (~78%) had received at least one previous scholarship: 13% had received one previous award, 48% had received two previous awards, and 17% had received three previous awards. Nearly all students were free/reduced lunch eligible (90%). The scholarship recipients primarily represented three racial/ethnic groups, Black/African American (65%), White/Caucasian (19%), and Hispanic (10%), and 6% were another race or no information was provided. Students resided in 51 counties in the state, with approximately 32% zoned to attend a failing school.

Students in grades kindergarten through second grade and grades 9 and 12 comprised 1445 (35%) of scholarship recipients and were not required to be tested according to the AAA because these grades are not tested in public schools in Alabama. A total of 2631 students were in grades 3 - 8, 10, and 11 (the grades aligned with state-mandated testing and included in this report). Test score reports were provided for 2246 (85%) of these students. Of the 385 students with missing test data, a small number (62) were absent or were not enrolled in the school at the time of testing and 8 students did not test because they were designated as special needs. For the remaining 315 students with missing data, no definitive explanation was provided.

Achievement Test Data for 2016-2017 Scholarship Participants

Twenty-one different standardized tests were given by 149 different schools. Unfortunately, some schools used tests that few or no other schools used. These schools typically had a low number of scholarship recipients. Making public these test results (especially when disaggregated by grade, race, or gender) would lead to undesirable results: a) Schools and individual children could be identifiable; the latter is a violation of FERPA; and b) Small samples, as noted earlier, are not likely to be representative of the full group of scholarship recipients. For these reasons, results from these schools would not contribute meaningfully to the AAA evaluation and therefore, the 255 students attending these schools were eliminated from this evaluation. Figure 1 provides a flow chart that summarizes factors affecting the 2016-2017 sample size.

Figure 1: Evaluation Sample Attrition



A total of 1,991 students or 76% of students for whom testing was required according to the AAA had potentially reportable test data from seven standardized tests: 1) The ACT (college entrance exam), 2) ACT Aspire (also used by ALSDE), 3) The Iowa Test of Basic Skills, 4) The PreACT (college entrance exam), 5) The Practice SAT-National Merit Scholarship Qualifying Test (PSAT/NMSQT), 6) The Stanford Achievement Test 10, and 7) Terra Nova. The table below indicates the number of students who took each test and the number of schools represented by each test. Students in this group attended 110 unique schools. The discrepancy between this total and the numbers listed in the table is due to some schools giving more than one test (e.g., a K-12 school might give the ACT Aspire for grades 3 - 8, the PSAT/NMSQT for grade 10, and the ACT for grade 11). Further attrition occurred due to missing test scores because schools might not have included a particular subject area in their reports, did not report usable scores (e.g., percent or number correct) or individual students may not have tested in a subject area. These instances are described as the results for each test are presented.

Table 1: Tests Included in the Evaluation for Grades 3 - 8, 10, and 11

Test	Number of Students	Number of Schools
ACT	46	21
ACT Aspire	331	20
Iowa Test of Basic Skills	1052	57
PreACT	78	6
PSAT/NMSQT	99	19
Stanford Achievement Test 10	343	26
Terra Nova	42	7
Total	**1,991**	

Description of Tests

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). They provide a score, such as a national percentile, that can be used to evaluate student performance relative to other students in the U.S. A child who scores at the 50th percentile is performing as well as or better than half of the students in the nation who are at the same grade level. Scale scores are derived from the number of items answered correctly and are often used to determine if students are meeting grade level benchmarks. Generally, scores on these tests are used to assess whether students or school systems have met requirements set by national or state standards, and consequently meet the testing requirement put forward in AAA. A brief description of each of the seven tests follows.

- The *ACT* is a nationally normed college entrance exam. The scores are used to predict college readiness. Reports include an ACT score (1-36) and a national percentile score. Subscale scores are provided for reading, English, and math. It is usually taken by high school juniors and seniors.
- The *ACT Aspire* assesses progress toward college and career readiness using criterion referenced benchmarks (e.g., state a definition, theorem, formula or axiom; understand cause and effect relationships; apply an understanding of the conventions of standard English grammar, usage, and mechanics to revise and edit text). The ACT Aspire includes test scores for reading, English, and mathematics, in addition to other areas. Scale scores indicate students' performance against a set of learning standards for each grade level. 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 also 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.
- *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. In contrast to the ACT Aspire benchmarks, interpreted 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 test has been validated at the national level, and it provides national percentile scores for reading, English, and math.
- *The Practice SAT - National Merit Scholarship Qualifying Test* (PSAT/NMSQT) is used to prepare students to take the SAT college entrance exam and is usually taken in the 10th and 11th grades of high school. The scores include a composite score that aligns with a predicted SAT score, as well as a subscale score in math and a combined reading and writing subscale score. National percentile scores are provided for all subject areas.
- *The PreACT* 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 college readiness indicators are provided for 10th graders. Subscale scores are provided for reading, English, and math. High school

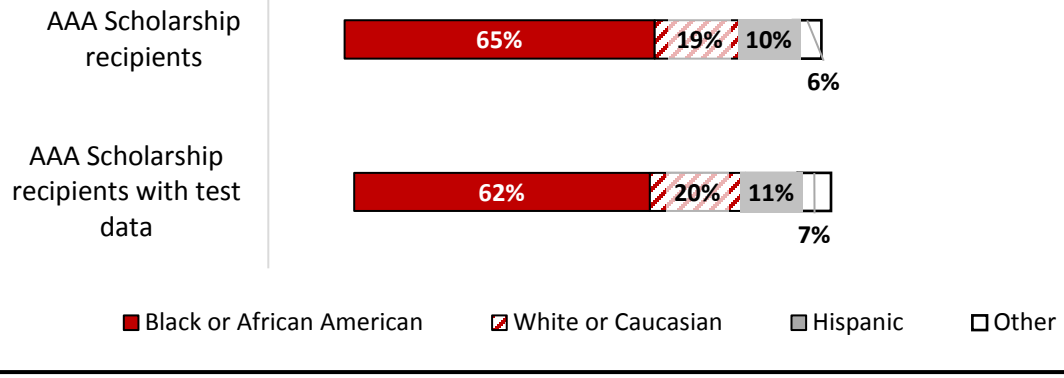
students commonly take this test their second year of high school. National percentile scores were not included in the test reports for the PreACT

- The *Stanford Achievement Test 10* is a norm-referenced test similar to the Iowa Test and was developed, among other reasons, to compare a child's academic achievement relative to others in the nation. The Stanford provides achievement/ability scores in language, reading, and math, including national percentile scores.
- *Terra Nova, 3rd edition* is a norm-referenced test similar to the Stanford Achievement Test and Iowa 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.

Demographic Information for Scholarship Recipients Included in the Evaluation

Based on information provided by the SGOs, the 1,991 scholarship recipients with usable test scores were very similar to the larger group of scholarship recipients. The majority had received a scholarship previously: Fifteen percent (15%) were first time scholarship recipients, 11% were two time scholarship recipients, 51% were three time recipients, and 22% were in their fourth year. (This information was missing for 1% of the students.) Nearly all were eligible for free or reduced lunch (90%). The SGOs reported that 34% (677) of the scholarship recipients were zoned to attend a public school that was designated as failing by the ALSDE. As with the larger sample (Chart 1), the racial/ethnic make-up of the sample was predominantly from three groups, Black/African American (62%), White/Caucasian (20%), and Hispanic (11%), and the remaining 7% of students were either another race, more than one race, or no race was designated. There were slightly more female (52%) than male students. Students represented 43 counties in the state and attended 114 different schools.

Chart 1: Racial Demograhics of the Evaluation Sample Compared to the Total Population of Scholarship Recipients



Findings for the 2016-2017 Academic Year

In this section, outcomes are described for each of the seven tests. For each test a brief description of the student demographics is provided, and additional test details relevant for understanding the test scores are given. When possible, test scores disaggregated by grade, race/ethnicity, and gender are presented. Statistical tests comparing scores among racial/ethnic groups and between genders were conducted when there were sufficient numbers of students in these groups ($n \geq 25$). National percentile scores are included for most tests. When relevant, scale scores were reported to aid in interpreting the test score information. Due to rounding, sometimes percentages in a table or chart sum to a number slightly greater or less than 100%.

Objective 1: Describe the Academic Achievement of Scholarship Recipients

The presentation of the results is organized by the type of test, norm- or criterion-referenced, since the tests within each type measure achievement in similar ways. The first three tests, Stanford Achievement Test 10, Terra Nova, and Iowa Test of Basic Skills are norm-referenced tests. The criterion-referenced ACT Aspire, PreACT, ACT, and PSAT/NMSQT are summarized next. The AAA asks for test scores 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 in this report as well. Due to the low representation of other races/ethnicities (typically 1.5% or less), descriptive information is only provided for Black/African American, White/Caucasian, and Hispanic groups.

Norm-Referenced Test Results

Stanford Achievement Test 10

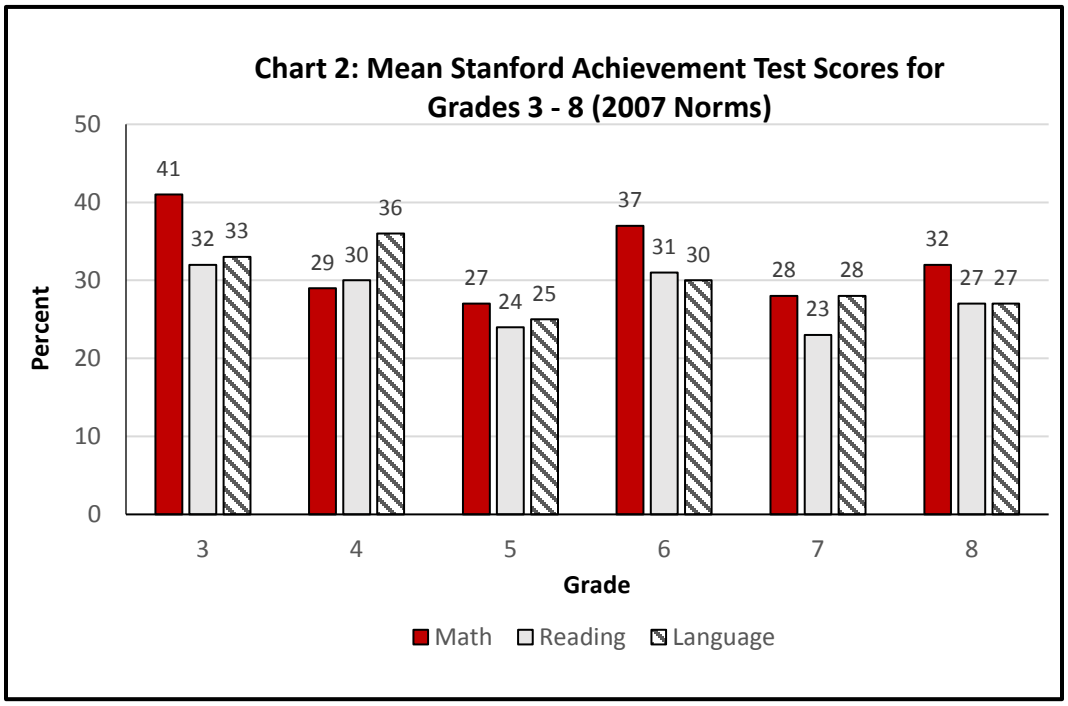
The Stanford Achievement Test 10 was given to 343 students in grades 3 through 11. There was not a sufficient number of test scores (25 or more) available for the 10th and 11th grades, which reduced the sample size to 312 for grades 3 through 8.

There were slightly more male than female students (52% male) and the racial make-up was predominantly Black/African American (72%), followed by White/Caucasian (22%). The majority of students had received a scholarship previously (82%), and most of this group (87%) had received two or more scholarships previously. As with the larger sample, nearly all students were free/reduced lunch eligible (93%).

As in previous years of this report, some schools used 2002 norms in reporting the national percentile scores, despite the availability of more recent 2007 norms. Although neither set of norms is current, the better indicator of student achievement relative to national standards is the more recent 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. Table 1A in the Appendix presents detailed data for each grade level for the sample as a whole and for demographic groups, with scale scores included as a reference point for the two sets of norms. Chart 2 below summarizes the findings for students who took the Stanford Achievement Test.

Examining the results for all children at each grade level revealed that the average percentile scores for each grade level were low, with the highest being 41% (3rd grade math) and the lowest being 23% for 7th grade reading as shown in Chart 2.

There were sufficient numbers of students (25 or more) to report scores separately for Black/African American, male, and female students for some grade levels. However, statistical comparisons could not be made among racial/ethnic groups due to the lack of a comparison racial group with a sufficient sample size. Gender comparisons could be made for 7th and 8th grade scores. These tests revealed that 8th grade girls performed significantly higher than boys in language arts (Means = 38th percentile vs. 15th percentile) and reading (Means = 36th percentile vs. 19th percentile. No other differences were significant.



Terra Nova

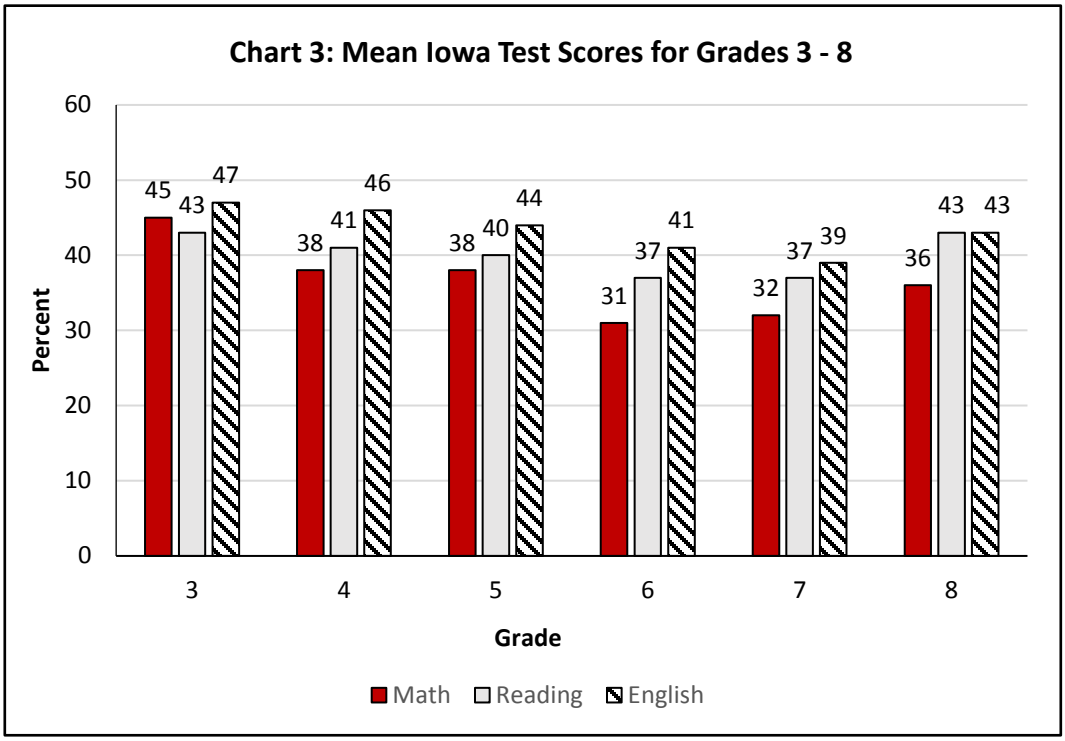
Usable Terra Nova test score results were available for 42 students in grades 3 through 8 who represented seven schools (Table 2). There were insufficient numbers of students at each grade level to report results disaggregated by grade, so the table below provides the results for all grades combined. The Terra Nova test takers were 62% female, 50% Black/African American, and 45% White/Caucasian. First time scholarship recipients comprised 19% of this group and 78% had been a scholarship recipient for three or more years. Nearly all of the students (91%) were free/reduced lunch eligible. Separate scores for racial groups are not reported due to the small numbers in each group. For gender, only female students had a sufficient number to meet the criterion for reporting. As a consequence of the low numbers, no statistical comparisons were made. Generally, the mean scores are at or slightly above 50%, suggesting that the students taking the Terra Nova ranked near the middle of students nation-wide who took the test.

Table 2: Mean Terra Nova Test Scores for Grades 3 – 8 Combined				
Grades	Group (N)	Math Percentile	Reading Percentile	Language Percentile
3 - 8	All (40-42)	50	55	54
	Female (24-26)	50	57	53

Iowa Test of Basic Skills

The Iowa Test was administered to 1052 students in grades 3 through 8 and in grades 10 and 11. However, not all test reports included the national percentile scores, reducing the Iowa test sample to 1032. The racial/ethnic make-up consisted of 58% Black/African American, 17% White/Caucasian, and 18% Hispanic students. First time scholarship recipients comprised 12% of the Iowa test takers, and 79% were in their third or fourth year of receiving a scholarship. The vast majority were free/reduced lunch eligible (93%), and 49% of the test takers were female.

The number of students in grades 10 (n = 11) and 11 (n = 6) were too small to provide reliable results. Chart 3 provides the results for the remaining grades and Table 2A in the Appendix provides the detailed data for each grade level disaggregated by gender and race when appropriate. Examining the results for all students at each grade level revealed that average scores ranged from 31% (6th grade math) to 47% (3rd grade English), indicating that on average scores were below the median for the country. There were large enough numbers to break down performance for every grade level by gender and race (White/Caucasian, Black/African American, and Hispanic). Statistical comparisons among the racial/ethnic groups were made for each grade level when there were sufficient numbers of students as indicated in Chart 4.



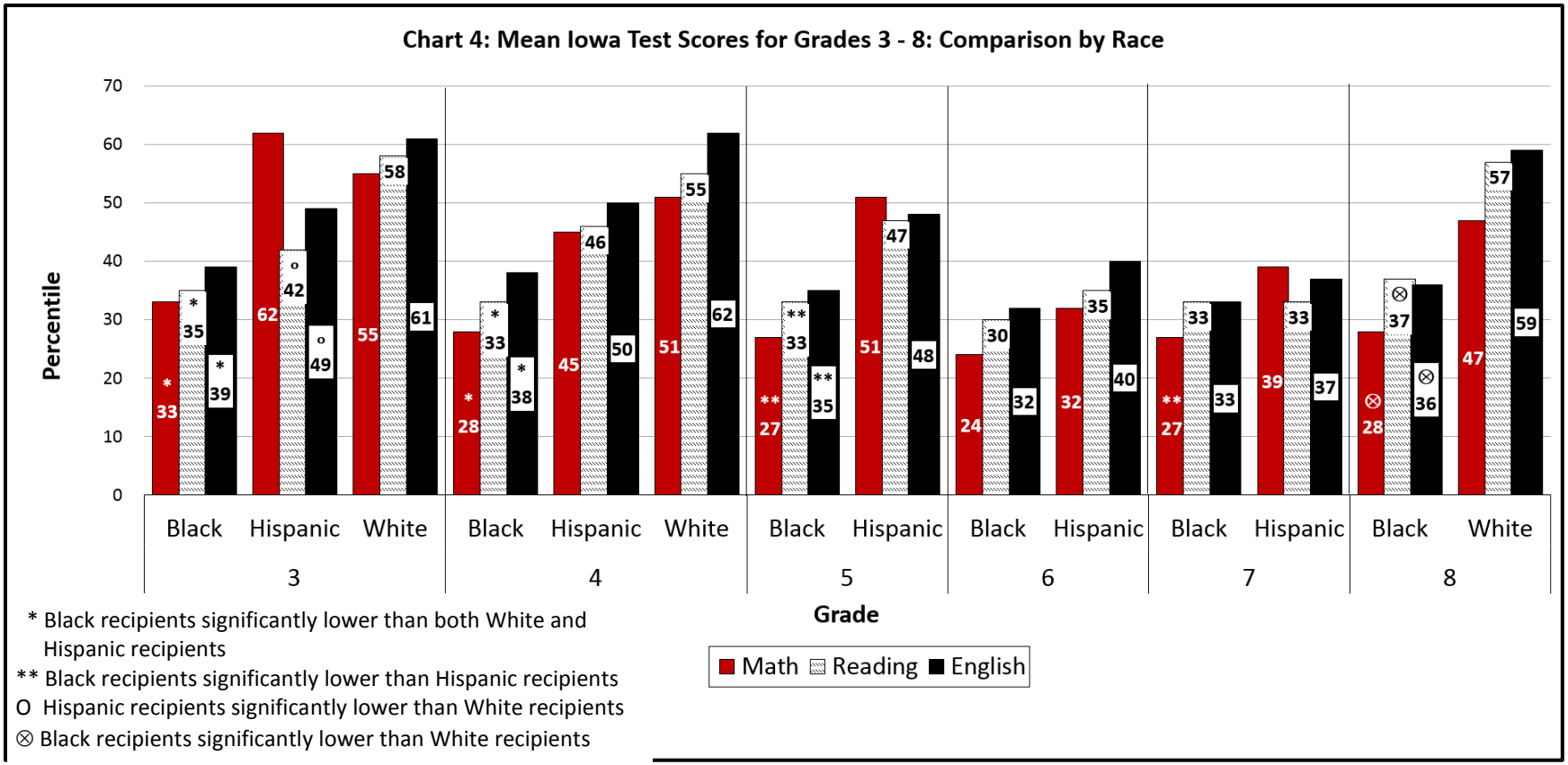


Chart 4: Mean Iowa Test Scores for Grades 3 - 8: Comparison by Race
Percentile
70
60
50
40
30
20
10
0
Black
Hispanic
White
3
4
5
6
7
8
Grade
Math
Reading
English
* Black recipients significantly lower than both White and Hispanic recipients
** Black recipients significantly lower than Hispanic recipients
O Hispanic recipients significantly lower than White recipients
⊗ Black recipients significantly lower than White recipients

Regardless of the subject area, White/Caucasian students scored significantly higher than Black/African American students. Hispanic students outperformed Black/African American students in all subject areas in grades 3 through 5, except 3rd grade reading. In 6th grade there were no significant differences between these two racial/ethnic groups, and in 7th grade Hispanic students scored higher in math. Only grades 3 and 4 had enough students to make comparisons between White/Caucasian and Hispanic students. For third graders, White/Caucasian students did better than Hispanic students in reading and English. There were no other differences between these two groups of students. Comparisons between genders for each grade level yielded one significant effect: In the fourth grade, girls scored significantly higher than boys in English (Means = 52nd percentile vs. 42nd percentile).

Summary for Norm-Referenced Test Results

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. Although the 50th 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 recipients' mean scores should be close to the 50th percentile, if as a group they are achieving at levels similar to others in the U.S. Generally, meeting or exceeding this standard would be considered a positive outcome. A review of the percentile scores provided in the previous three tables indicates that nearly all of the average percentile scores were below the 50th percentile. Subsequently, statistical comparisons were made separately for each test to assess if the mean scores in each subject area were significantly lower than the 50th percentile.

Considering first the Stanford Achievement Test, the average scores for each grade level and subject area were generally significantly below the 50th percentile. The exceptions were for 3rd grade math (all 3rd graders, African American/Black, and females) and language for female students, which were not significantly different from the 50th percentile score.

For the Iowa Test, with one exception, the mean scores for all students (combined across race and gender) at each grade level were significantly below the 50th percentile mark. The exception was for 3rd grade English, which was not significantly different from the 50th percentile. The results differed somewhat for the three racial/ethnic groups.

- Black/African American students performed significantly lower than the 50th percentile at all grade levels.
- White/Caucasian students' scores were significantly above the 50th percentile for 3rd grade reading and English and for 4th grade English, but did not differ from the 50th percentile in any other comparison.
- Hispanic students' scores were significantly higher than the 50th percentile for third grade reading, but lower than the 50th percentile for nearly all subjects in 6th and 7th grade. Hispanic students' scores did not significantly differ from the 50th percentile in all other grades and subjects.

Male and female students' scores were generally significantly below the 50th percentile scores. Exceptions in which scores were not significantly lower were for 3rd, 4th, and 5th grade English for girls and 4th grade math for boys.

For the Terra Nova tests, due to the small number of students, mean percentile scores were combined across all grade levels as indicated previously. The mean percentile scores combined across all grade levels did not significantly differ from the 50th percentile. This replicated the findings from the previous years and suggests a generally better outcome for this group of students. For three years in a row, students taking the Terra Nova test did not perform significantly lower than the students in the U.S., performing at the median. In previous reports several hypotheses were entertained to explain why the results for Terra Nova were relatively better compared to the other tests, which are applicable for this report as well. Discrepancies could be due to differences in the content of the tests or differences in the schools that happened to choose one test over another that may produce higher levels of achievement (e.g., curriculum, pedagogical approaches). The demographic make-up of this group of test takers was somewhat different than those who took the Iowa and Stanford tests, having relatively more White/Caucasian and female students, which are both factors that are sometimes associated with higher standardized test scores. However, the free-reduced lunch rate was similar. There were only 42 students included in this group compared to 1344 students who took either the Iowa or the Stanford. Thus, due to small sample size, the results from the Terra Nova are likely not representative of the larger group of scholarship recipients.

The summary graphic below provides the key findings for the norm-referenced tests. Together the results generally indicate that the scholarship students as a group do not perform better than the national average on these tests, and it is more typical for them to perform below average compared to other students in the U.S. In each of the previous reports there have been anomalous findings to this generalization for specific grades and standardized tests, as there are in this report. No discernable pattern has emerged that would suggest that the grade level of the scholarship recipient impacts performance on the standardized tests in a systematic way. As noted in previous reports, there is likely variation among schools, tests, and students that cannot be accounted for in this evaluation. With the relatively small sample sizes there is an increased probability of variation among the grades that may result in some grades performing better than others. This could be due to any number of performance related factors, such as ability, having a good testing day, or difference in teacher quality, among others.

When it was possible to make racial/ethnic comparisons using the larger sample of Iowa test scores, Black/African American students performed more poorly than White/Caucasian students, and although less consistently found, they also performed more poorly than Hispanic students. Hispanic and White/Caucasian students could only be compared in two grades where the groups did not differ in four of the six comparisons made.

Summary for Norm-Referenced Test Results

Scholarship students as a group did not perform better than other students in the U.S.

- It was most typical for students to perform below average compared to the nation.
- Outcomes were even poorer for African-American participants who made up the majority of scholarship recipients (65%).

There are anomalous findings to this generalization for specific grades and standardized tests:

- No discernable pattern has emerged indicating that the grade level of the scholarship recipient impacts performance on the standardized tests in a systematic way.
- Small sample sizes for some grade levels and tests adversely impact the reliability of these findings.

Stanford Achievement Test	Terra Nova	Iowa Test
The average scores for each grade level and subject area were at or significantly below the 50th percentile. **Exceptions:** ✓ 3rd grade math ✓ 3rd grade language arts for females	The mean percentile scores combined across all grade levels were not significantly different from the 50th percentile.	The average scores for each grade level and subject area were generally significantly below the 50th percentile. **Results by racial group:** ✓ African-American students scored significantly lower than White and Hispanic students in some subjects and grades and below the 50th percentile overall ✓ Third grade White students did better than Hispanic students in reading and English ✓ Hispanic and White students scored above the 50th percentile for some subjects in 3rd and 4th grade

Criterion Referenced Test Results

ACT Aspire

The ACT Aspire test was administered to 331 students (61% female) in grades 3 through 8 and 10. Similar to other test groups a minority of students (27%) were first time scholarship recipients. Students who took the ACT Aspire were 61% Black/African American, 24% White/Caucasian, and 9% Hispanic. Most were eligible for free/reduced lunch subsidies (80%).

As noted earlier, scoring by ALSDE identifies four proficiency benchmarks that classify students as 1) *In need of support*, 2) *Close*, 3) *Ready*, and 4) *Exceeding* for reading and math. Students who are at or above the benchmark level 3 are considered to be on track to be college ready by 11th grade. Although not publicly reported for public school students, ALSDE does provide benchmarks for English. The average scale score, its corresponding national percentile rank and proficiency level are provided for each grade level in Table 3. There were sufficient numbers of students in many grade levels to disaggregate scores for Black/African American students and females. Table 3 below presents ACT Aspire scores for these groups when the sample size met the minimum standard.

The ACT Aspire is a criterion-referenced test, and as a result, the national percentile scores are interpreted differently compared to a norm-referenced test. For example, for reading, the average national percentile scores ranged from 47% to 56%, scores that are generally higher than those evident for the norm-referenced tests presented earlier. However, only one grade (grade 8) had average scores that fell in the "ready" proficiency category for reading. Because ALSDE only reports the proficiency group results, these ratings are the focus of this report.

Examining the results presented in Table 3, there is a decidedly different pattern of results for English compared to math and reading. For all grade levels and subgroups, the average English scores reached the benchmark for "ready" or "exceeding." In contrast, the majority of grade level averages for math and reading fell below benchmarks. The exceptions were for 3rd, 4th and 6th grade math and 8th grade reading, all of which reached the "ready" benchmark. African American/Black students reached the "ready" proficiency level for math in grades 3 and 7, above the level for the grade as a whole. In contrast, they did not reach the "ready" benchmark at any grade level for reading. Female students' scores evidenced the same levels of proficiency on average as the sample as a whole at each grade level.

Because the ACT Aspire is also used by the ALSDE it was important to provide additional detail for this group of students. For each grade level, Charts 5-7 present the percentage of students at each proficiency level. (Detailed information on proficiency group percentages for each grade level is available in Table 3A in the Appendix.) Each subject area presents a different pattern of findings.

Table 3: ACT Aspire Scores for Grades 3 – 8 and 10

Grades	Group (N)	Math			Reading			English		
		SS	%	Prof.[1]	SS	%	Prof.[1]	SS	%	Prof.[1]
3	All (54)	413	47	3	412	50	2	417	53	3
	Black (40)	413	47	3	412	51	2	418	56	4
	Female (31)	412	46	2	412	50	2	417	54	3
4	All (39-40)	416	52	3	415	54	2	421	57	3
	Black (27)	415	49	2	415	50	2	421	55	3
	Female (*)	*	*	*	*	*	*	*	*	*
5	All (37-38)	417	45	2	416	47	2	423	50	3
	Black (*)	*	*	*	*	*	*	*	*	*
	Female (26-27)	417	46	2	417	50	2	423	53	3
6	All (37-40)	423	61	3	419	51	2	427	58	4
	Black (*)	*	*	*	*	*	*	*	*	*
	Female (28)	423	61	3	420	55	2	427	61	4
7	All (54-57)	420	52	2	420	53	2	428	56	3
	Black (32)	422	58	3	421	55	2	428	56	3
	Female (35)	419	48	2	421	53	2	428	56	3
8	All (54-57)	422	50	2	424	56	3	430	59	4
	Black (27-28)	421	46	2	422	49	2	428	54	3
	Female (33)	421	48	2	424	53	3	429	57	4
10	All (39-43)	422	37	1	422	47	2	429	45	3
	Black (26-29)	421	35	1	421	45	1	428	43	3
	Female (≤25)	*	*	*	423	50	2	431	49	3

* Indicates an insufficient number of students in the group (< 25) for reporting.
[1] Prof. = Proficiency Groups: 1—*In need of support*, 2—*Close,* 3—*Ready*, and 4—*Exceeding*
SS = Scale Score, % = percentile score

For math (Chart 5), the majority of students in grades 3, 4, and 6 scored at level 3 (Ready) or Level 4 (Exceeding), but at all other grade levels the majority of students did not reach proficiency benchmarks. In 5th grade half of students are at Level 2 (Close), and in the 10th grade, most students performed at Level 1 (In need of support). In contrast, the percentage of students meeting benchmarks for reading was much lower (Chart 6). The majority of scholarship recipients were in proficiency levels 1 or 2 for most grades. Only the 8th grade had a majority of students reaching proficiency levels 3 and 4. Finally, the majority of students at all grade levels performed at levels 3 or 4 in English in grades 3 - 8 (Chart7). In 10th grade, about half of students performed at levels 3 or 4.

Chart 5: ACT Aspire Proficiency Groups for Grades 3 – 8 and 10 for Math

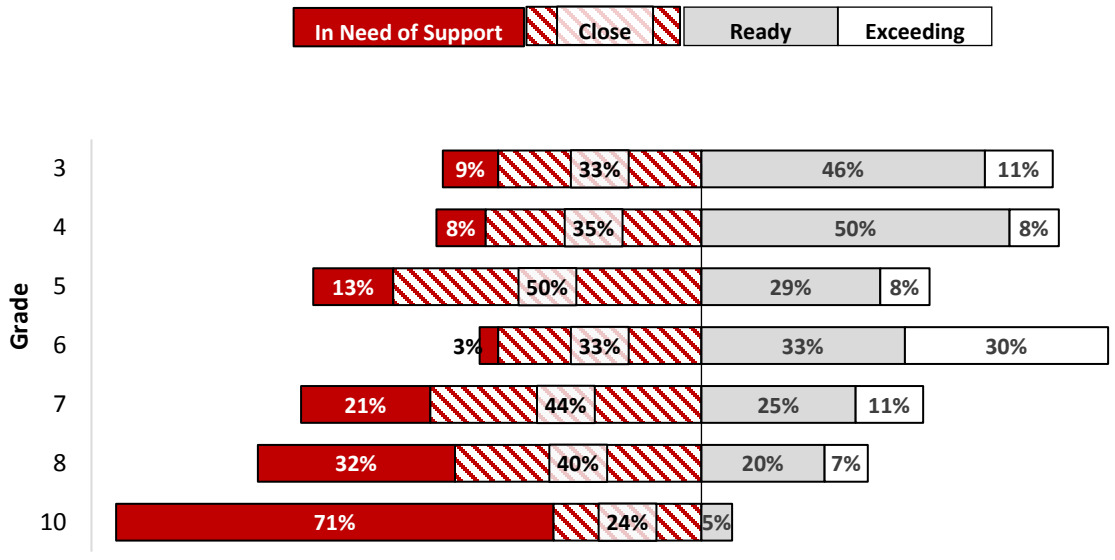
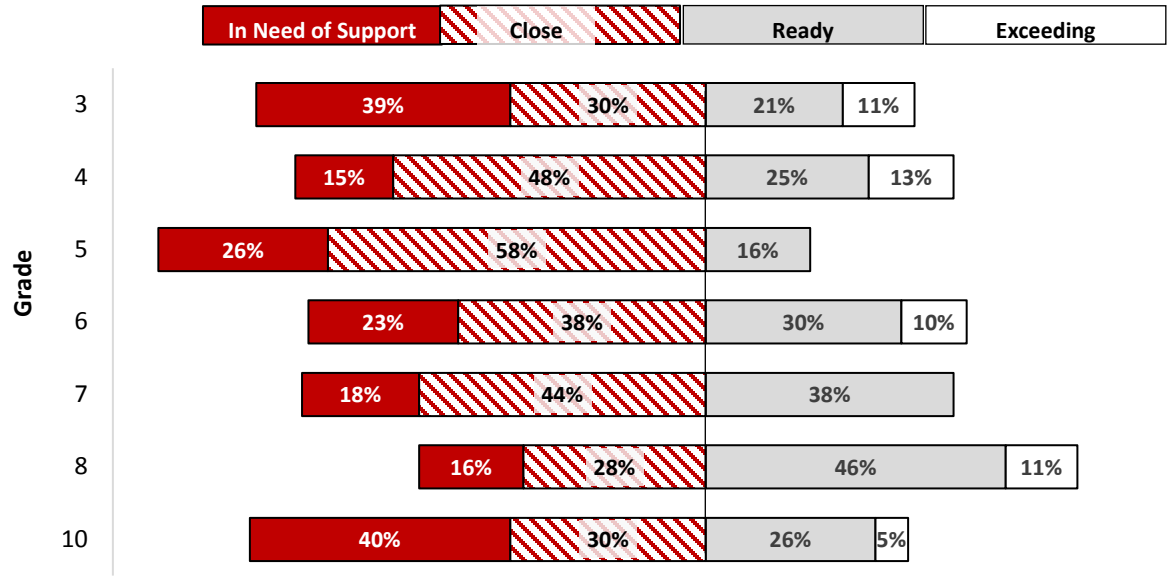
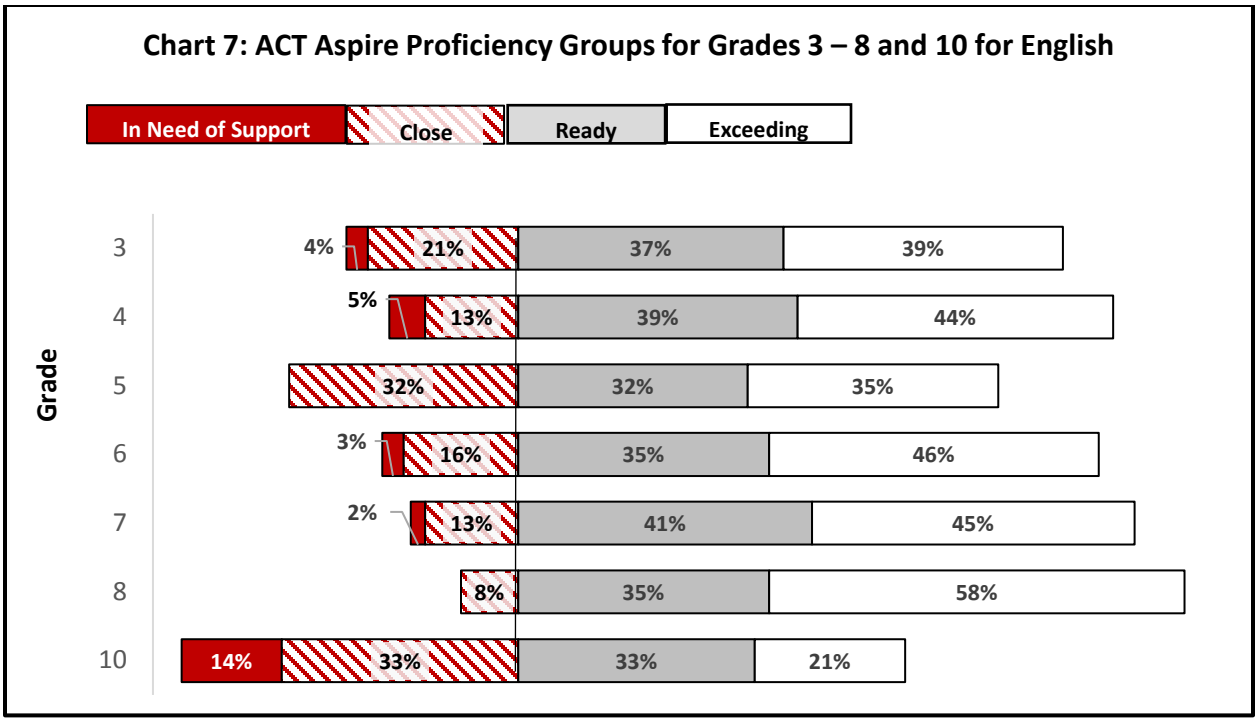


Chart 6: ACT Aspire Proficiency Groups for Grades 3 – 8 and 10 for Reading





PSAT/NMSQT

The PSAT/NMSQT was administered to 99 students in grades 8, 10, and 11. Of these students, only 4 (4%) were first time scholarship recipients, 10% had received their first scholarship the previous academic year (2015-2016), and the remainder (86%) had participated in the scholarship program for three years or more. These students were nearly evenly divided by gender (49% female), and nearly all were free/reduced lunch eligible (90%). The racial/ethnic make-up was 74% Black/African American, 17% White/Caucasian, and 5% Hispanic. Only grades 10 and 11 met the minimum requirement of 25 students for reporting (43 and 48 students, respectively). Only the Black/African American racial/ethnic group was represented in sufficient numbers to report their results separately, and neither gender met the sample size criterion to report disaggregated results. The PSAT/NMSQT combines reading, writing, and language scores into an "evidenced-based reading and writing score." As a result, the combined percentile scores are presented in Table 4.

The reading-writing and the math scores are aligned with benchmarks used to predict college readiness. The benchmark scores correspond to a 75% likelihood of achieving a grade of "C" or better in the first semester of college for courses in related areas. Scoring for the PSAT/NMSQT places students' scores into one of three categories: *Need to strengthen skills*, A*pproaching benchmark*, or *Met or exceeded benchmark*. For both 10th and 11th grade students, the average math score is categorized as "Need to strengthen skills." For 11th graders the average reading-writing score met the benchmark, but for 10th grade this score fell into the "approaching benchmark" range. The average scores for African American/Black students in 11th grade were categorized as "need to strengthen" for both reading and math.

Table 4: Mean PSAT/NMSQT Scores for Grades 10 and 11					
Grade	Group (N)	Math		Reading-Writing	
		Percentile	Benchmark	Percentile	Benchmark
10	All (40-41)	27	Need to Strengthen	34	Approaching
	Black (28-29)	23	Need to Strengthen	31	Need to strengthen
11	All (33-36)	28	Need to strengthen	42	Met
	Black (*)	*		*	
* Indicates an insufficient number of students in the group (< 25) for reporting.					

Examining the full distribution of scores for each grade level revealed the following patterns:

- 10th grade: 15% met the math benchmark; 36% met the reading-writing benchmark
- 11th grade: 15% met the math benchmark; 42% met the reading-writing benchmark

PreACT Test

The PreACT Test was administered to 78 students in grades 10 and 11. The racial/ethnic make-up of this group of students was 60% Black/African American, 30% White/Caucasian, and 1% Hispanic. Females (55%) had a slight majority over males (45%), and similar to the larger sample, most (90%) were free/reduced lunch eligible. Only 6% were first-time scholarship recipients, 12% had received one previous award, and 82% had received a scholarship for three or more years. For 10th grade there were sufficient numbers of students to disaggregate scores for Black/African Americans and female students (Table 5).

The PreACT reports did not include percentile scores, only scale scores (range 1-36) that correspond to the ACT college entrance exam scores. Benchmark scores are provided to indicate college readiness. Specifically, these benchmarks indicate, "the level of achievement required for students to have a 50% change of obtaining a B or higher or about a 75% chance of receiving a C or higher in corresponding credit-bearing first-year college courses" (PreACT Technical Bulletin p. 32). These benchmark scores correspond to the same benchmarks adopted by ALSDE: 22 for math, 22 for reading, and 18 for English.

Because the ACT is normally taken in the 11th grade, additional college readiness indicators are provided for 10th graders. The rationale behind the additional indicators is that 10th grade students will continue to gain skills and knowledge over the course of the year. As a result, these indicators can be used to make predictions as to the likelihood of meeting the benchmark scores in 11th grade. Three benchmark levels are defined for each subject area: *In need of intervention, On the cusp,* and *On target.* Table 5 presents the mean scores for 10th and 11th grade students and provides the corresponding college readiness indicator level for 10th graders.

Table 5: Mean PreACT Test Scores for Grades 10 and 11							
Grade	Group (N)	Math		Reading		English	
		Scale Score	Readiness Indicator[1]	Scale Score	Readiness Indicator[1]	Scale Score	Readiness Indicator[1]
10	All (52-53)	16	Intervention	18	On Cusp	15	On Target
	AA/Black (32-33)	15	Intervention	16	Intervention	14	On cusp
	Female	17	On Cusp	18	On Cusp	16	On Target
11	All (25)	17	NA	18	NA	17	NA
	AA/Black (*)	*		*		*	
	Female (*)	*		*		*	

[1] Readiness indicators are for 10th grade students only. NA = not applicable

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

Note. College benchmark scores are 22 for math and reading and 18 for English.

For 11th graders, the mean scale scores fell below benchmarks for college preparedness. To further investigate, the percentages of the students who met or exceeded the benchmark scores were calculated and indicated rates of 12% for math, 24% for reading, and 48% for English.

With the exception of English, the 10th grade scores generally did not meet the readiness benchmark. The percentages of students who fell into each of the three readiness categories were calculated, and the results are presented in Table 6.

Table 6: Percentage of Students in Grade 10 within each Readiness Category for the PreACT									
Grade (N)	Math			Reading			English		
	Inter-vention	On Cusp	On Target	Inter-vention	On Cusp	On Target	Inter-vention	On Cusp	On Target
10 (52-53)	66%	25%	10%	57%	15%	29%	27%	15%	58%

Together these results suggest that the majority of 10th grade scholarship recipients who took the PreACT Test failed to meet national standards predictive of college achievement in math and reading. However their performance in English was much better, with more than half meeting the criterion of *on target.*

ACT

The ACT was administered to 46 students in grades 10 and 11. The majority of this sample was Black/African American (63%), followed by 15% White/Caucasian and 9% Hispanic. All but one student was eligible for free/reduced lunch and 52% were female. Similar to the other tests, only a small percentage were first year scholarship recipients (15%) and the majority (74%) were in their 3rd or 4th year of receiving a scholarship. Only the 11th grade had a sufficient number of students (32) to report scores, but there were not enough students to break out these scores by gender or race. Both percentile scores and scale scores are presented in the table below. The ACT College Entrance Exam scale scores range from 1 to 36.

Table 7: Mean ACT Scores for Grade 11							
Grade	Group (N)	Math		Reading		English	
		Scale Score	Percentile	Scale Score	Percentile	Scale Score	Percentile
11	All (32)	17	33	18	38	18	40

ALSDE has set benchmark scores for 11th grade ACT scores, which are identical to those described for the PreAct (22 for Math, 22 for Reading, and 18 for English) and interpreted similarly. The average ACT scale scores fell below benchmark scores for college preparedness for reading and math, but met the benchmark for English.

Additionally, four proficiency groups are identified by ALSDE that are comparable to those for the ACT Aspire. The percentages of 11th grade students falling into each group are presented below.

Table 8: ACT Proficiency Groups for Grade 11												
Group (N)	Math Proficiency Groups				Reading Proficiency Groups				English Proficiency Groups			
	1	2	3	4	1	2	3	4	1	2	3	4
All (32)	72%	6%	16%	6%	44%	38%	12%	6%	38%	16%	9%	38%
Proficiency Groups: 1—*In need of support*, 2—*Close,* 3—*Ready*, and 4—*Exceeding*												

The percentages of the students in 11th grade who met or exceeded the benchmark scores were 22% for math, 18% for reading, and 47% for English. Together these results suggest that the scholarship recipients who took the ACT generally failed to meet national standards predictive of college achievement, although performance was better in English, relative to other subjects.

Summary for Criterion-Referenced Test Results

The key performance indicator for students taking criterion-referenced tests is the number of students making benchmarks on each of the tests. The summary graphic below presents the principal findings.

For students in grades 3 through 8, only the ACT Aspire findings are applicable, and results varied depending on the subject area. The majority of students at all grades and demographic groups achieved scores that met or exceeded benchmarks for English, but for reading, only 8th graders achieved this level of proficiency on average. The results for math fell between these two extremes, with three grade levels (3, 4, and 6) having the majority of students meeting or exceeding proficiency levels.

Tenth graders were represented in three different tests. The results were similar for the ACT Aspire and PreACT in that the majority of 10th graders met benchmarks for English, but not in any other subject area, showing the weakest performance in math (only 5% to 10% meeting proficiency

score targets). The majority of 10th graders did not meet benchmarks on the PSAT/NMSQT for reading-writing or math.

Eleventh grade students were also represented in three standardized tests: the PSAT/NMSQT, the PreACT or the ACT. Findings were similar across the three tests in that the majority of 11th grade students did not meet benchmark scores in math, reading, or English.

Taken together, the pattern of results suggests that most of the scholarship students did not make proficiency benchmarks in math and reading. In these subject areas, there were some grades that defied this trend, notably in math for grades 3, 4, and 6 and in reading for grade 8. However, the majority of high school students on each of the four achievement tests performed below the math and reading benchmarks. For English, students across grade levels and tests generally performed better relative to math and reading. This was especially true for the ACT Aspire test where the majority of students met the benchmarks for English. It is not clear why English scores are generally better than math and reading, but it is a bright spot in this report.

Summary for Criterion-Referenced Test Results
Generally scholarship recipients did NOT meet benchmarks for reading and math, although there were some exceptions:
✓ The majority of scholarship recipients in the **3rd, 4th**, and **6th** grade **Met** or **Exceeded** benchmarks in **Math** on the **ACT Aspire**. ✓ The majority of scholarship recipients in the **8th** grade **Met** or **Exceeded** the benchmark for **Reading** on the **ACT Aspire.**
For English, the pattern of findings varied by achievement test and grade level:
✓ On the **ACT Aspire,** in grades **3 - 8 and 10 the majority of** scholarship recipients generally **Met** the benchmarks for **English.** ✓ In grades **10 and 11,** the majority of scholarship recipients **failed** to meet benchmarks for **English** on the **PSAT/NMSQT** and **ACT.**

Objective 1 Conclusion

Together the results generally indicate that the scholarship students as a group did not meet national achievement norms or benchmarks, although some exceptions have been noted. These findings need to be placed in a larger social context. The students in the AAA program belong to demographic groups (low income, racial minority groups) that have traditionally lagged behind other students in the state and in the country in academic achievement. Furthermore, many indicators, such as the NAEP assessments and the ALSDE annual reports, indicate that students attending public schools in the state of Alabama fall short of national standards. The information presented so far does not indicate whether the scholarship recipients' academic achievement represents an improvement, decline, or no change as a result of the AAA, nor does it indicate how these students directly compare to public school children in the state of Alabama. The next section of the report provides some insights on these issues.

Objective 2: Compare Scholarship Recipients to Alabama Public School Students

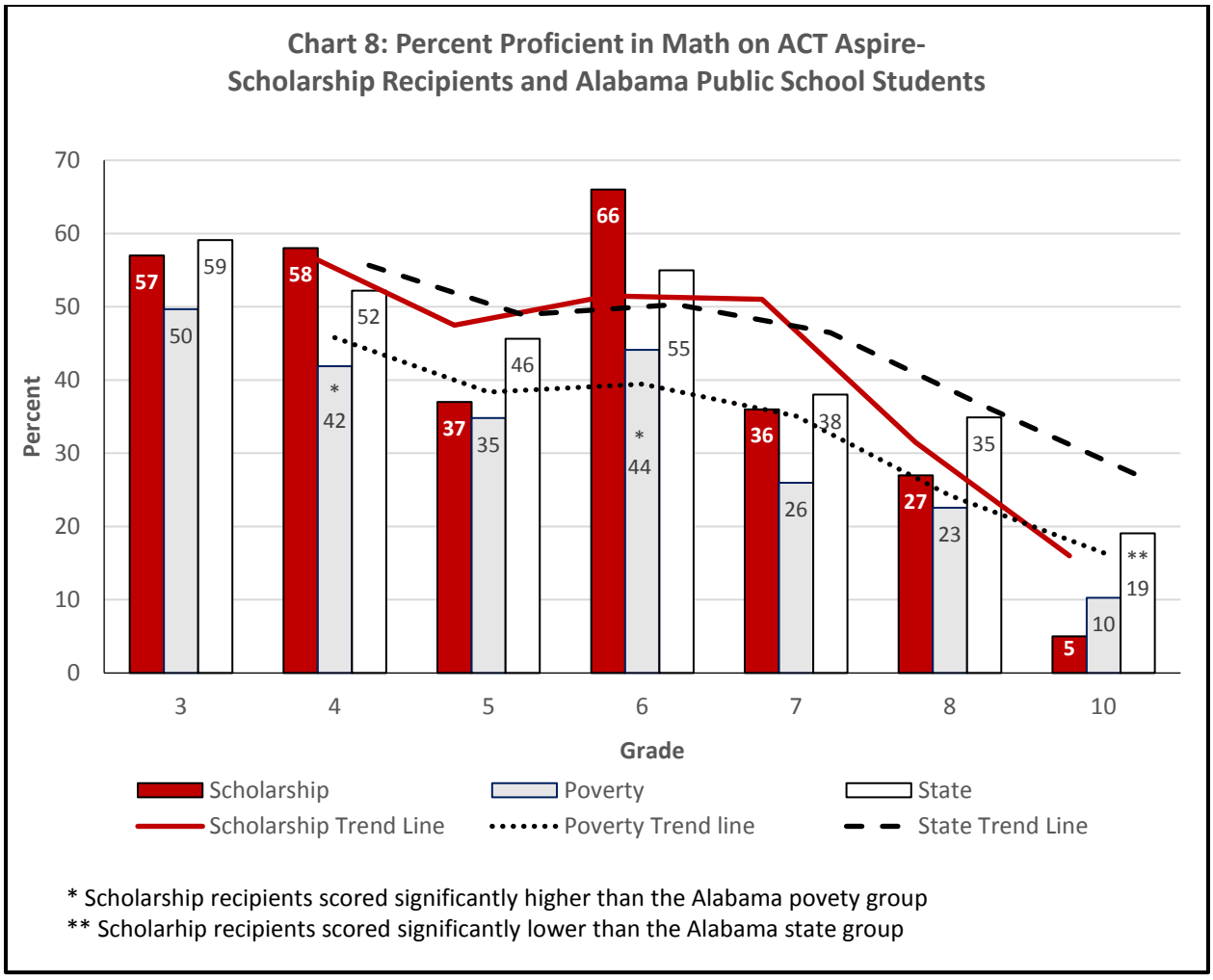
As in previous reports, the 2016-2017 scholarship recipients' performance on the ACT Aspire (grades 3 - 8 and 10) and the ACT college entrance exam (11th grade) are compared to the performance of Alabama public school students in general and Alabama public school students who receive reduced or free lunch (AL poverty). This comparison provides a snapshot of how the scholarship recipients perform in comparison to others in the state attending public schools without a scholarship.

As previously noted, the comparison between scholarship recipients and students attending public schools focuses on ACT Aspire proficiency scores due to the limited availability of ALSDE data. There are some significant limitations related to the interpretation of the results that must be noted. This small subsample of students, approximately 14% of scholarship students in the grades required to be tested, may not be representative of all of the participants in the AAA program. The relatively small number of scholarship students with ACT Aspire scores (331 in grades 3 - 8 and 10) and ACT college entrance exam scores (32 11th graders) represents only 30 (22%) of the 131 schools that students attended. There may be factors associated with the schools that used ACT Aspire and the ACT (as opposed to other tests such as the Iowa) that make these schools unrepresentative of the rest of the schools with scholarship recipients (e.g., demographic characteristics of students, class sizes, teacher quality, and pedagogical approaches). For example, the percentage of Hispanic students among those who took the Iowa test (18%) is twice the percentage among those who took the ACT Aspire (9%). With these limitations in mind, the comparisons that are set forth in the evaluation requirements for the AAA were made.

Proficiency Rates of Scholarship Students Compared to Alabama Public School Students

ACT Aspire

To compare scholarship students to public school students, a proficiency rate (percentage achieving at levels 3 and 4) was calculated for three groups of students for whom ACT Aspire scores were available: scholarship students, Alabama state poverty students (those receiving a free or reduced lunch subsidy) and all students in the state attending public schools. The percentage of children achieving proficiency at each grade level is represented graphically in Charts 8 (math) and 9 (reading). Table 4A in the Appendix shows the rates across all four proficiency groups. Small discrepancies between table and values in the charts that follow are due to rounding.) The poverty group provides the most appropriate comparison group, as nearly all scholarship recipients are eligible for the subsidy. Statistical analyses compared the proportion of scholarship recipients who achieved or exceeded proficiency at each grade level and subject area to the comparable scores for those who attended public schools. In Charts 8 and 9 asterisks within the bar for Alabama Poverty (*) or State (**) (which includes all demographic groups combined) indicates that there was a significant difference between the percentage of students who were proficient in that group compared to the scholarship recipients. The average moving trend lines are included in the charts to show the average change in percentile scores between each grade level for each comparison group. Average moving trend lines allow patterns in data to be visualized more clearly when there are fluctuations in the data. However, it is important to note that data presented at each grade level are from distinct groups of students and do not represent changes in individual students over time.

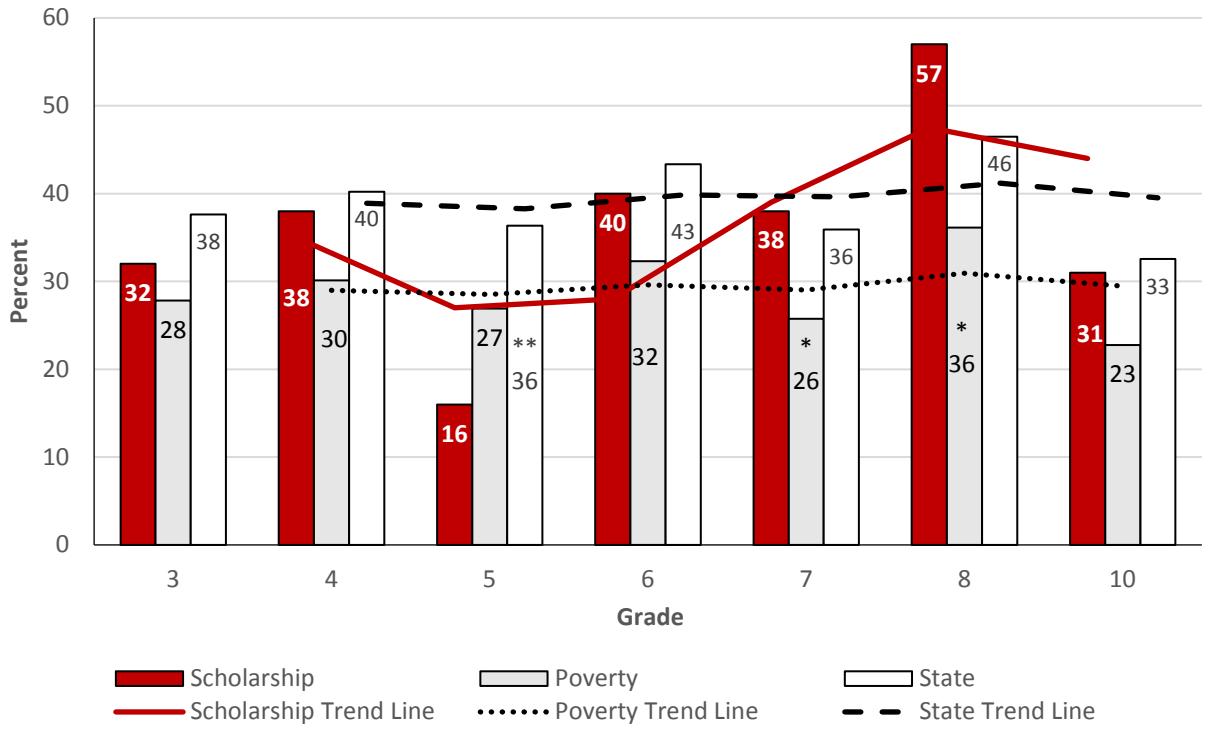


Comparing the scholarship recipients to Alabama public school students and the Alabama Poverty group revealed a mixed pattern of results that differed across the two subject areas. As reported in previous years, the percentage of students who are proficient in math decreased as grade level increased, regardless of scholarship status. As can be seen in Chart 8 by the 10th grade less than 20% of students are proficient in math across all three comparison groups. There were no significant differences between the scholarship students and the public school students for grades 3, 5, 7, and 8. However, in grades 4 and 6 the scholarship students' scores were significantly higher than the Alabama Poverty group (but not the state as a whole), and in the10th grade, scholarship recipients were less proficient in math than their counterparts in the state as a whole.

With respect to 2016-2017 reading achievement, regardless of scholarship status, generally less than 50% of students in any group were proficient in reading (Chart 9). Eighth grade scholarship recipients were the exception to this generalization, with more than 50% of students meeting proficiency benchmarks for reading. There were no significant differences between the three groups for grades 3, 4, 6, and 10. In grades 7 and 8 scholarship recipients scored higher than the Alabama Poverty group, but not the state as a whole, scholarship students and in 5th grade scholarship recipients were significantly lower than 5th grade students in public schools for the state as a whole. It is important to note that there appear to be large differences between some of

the scores that are not listed as statistically significant. Given the small sample sizes and the variability in the scores, the statistical testing tells us that we cannot be certain that these differences are not an artifact of the samples, rather than a true difference that would generalize to all scholarship students.

Chart 9: Percent Proficient in Reading on ACT Aspire- Scholarship Students and Alabama Public



*Scholarship recipients scored significantly higher than Alabama poverty group
** Scholarship recipients scored significantly lower than Alabama state group

ACT

For 11th graders taking the ACT, the ALSDE assessed college and career readiness by determining the percentage of students who met at least one benchmark in any subject area (English, math, reading, and science). For the state as a whole, approximately 50% met this standard for 2016-2017. This is comparable to the rate for the scholarship students, which was also 50%. However, it should be noted that science scores were not available for the scholarship students, so 50% may underestimate the number of students meeting this standard.

Additional data available from ACT provides the percentage of students meeting benchmarks for each state and the U.S. as a whole. ACT reports that the 2017 Alabama state-wide percentages of students meeting the benchmarks were 23% for math, 36% for reading, and 52% for English. For scholarship students the percentage of students meeting benchmarks for math (22%) and English (47%) were comparable, but the percentage making the reading benchmark (18%) was significantly lower than the Alabama-state-wide results. For all students in the U.S., 41% met the benchmark for math, 47% for reading, and 61% for English. These rates are statistically higher than those observed for the scholarship students for math and reading, but not English.

Objective 2 Conclusion

With a few exceptions noted above, comparisons between the scholarship students and the students attending public schools in Alabama generally indicate that the two groups continue to fall short of meeting the benchmarks on standardized tests. 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, and the small percentage of scholarship students included in this analysis is a concern. Additional issues 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.

The comparisons at each grade level did not present a cohesive pattern. In some cases, scholarship students performed better than the Alabama poverty group (for math in 4th and 6th grades and for reading in 7th and 8th grades), but in other cases they performed significantly worse than the state as a whole (5th and 11th grade reading, and 10th grade math). Scholarship students in 11th grade taking the ACT were comparable to the Alabama public school students on most indicators, only differing from public school students in meeting reading benchmarks, where fewer scholarship students were proficient.

To summarize, for both groups of students, there were very few instances where the percentage of students reaching proficiency was 50% or higher, suggesting there is need for improvement in the state as a whole. It is noteworthy that in 25 out of the 34 (78%) comparisons made there was no significant difference between the scholarship recipients and students attending public schools in the state. In cases where differences were observed, no reliable patterns across grade levels and subject could be discerned. As a result, there is no compelling evidence to suggest that on the ACT Aspire and ACT tests the scholarship students perform better or worse than public school children in Alabama.

Summary for Objective 2: Scholarship Recipients vs. Alabama Public School Students

- **Comparisons did not present a clear pattern across subjects and grade levels indicating that one group performed better or worse than the other.**
- **Scholarship students in 11th grade generally performed similarly to their public school counterparts.**

ACT Aspire findings for each grade level

✓ For **math**, in grades **4** and **6** the scholarship students' scores were significantly **higher** than the **Alabama Poverty** group, and in **10th** grade, scholarship recipients were **less** proficient in math than their counterparts **in the state as a whole**. For **all other grades** there were **no differences in** math in achievement.

✓ For **reading**, in grades **7** and **8** scholarship recipients scored higher than the **Alabama Poverty** group, and in **5th** grade scholarship recipients were significantly **lower** than their counterparts in public schools for the **state as a whole**. There were **no significant differences in any other grade.**

ACT findings for 11th graders

✓ Scholarship students in **11th** grade taking the ACT differed from public school students in meeting **reading** benchmarks, where **fewer scholarship students were proficient**, but they were **on par** with public school students **on other indicators.**

Objective 3: Changes in Achievement across Time

The third objective of this report addresses the question as to whether participation in the scholarship program over time results in achievement score changes that meet, exceed, or fall behind those of students attending public schools. Several approaches were taken to assess how the scholarship students' scores changed over time and how that change compares to students attending public schools. Ideally, such an analysis would calculate the average change in national percentile scores or proficiency groups over time for scholarship students and public school students, and then comparisons would be made between the two groups of students taking into account grade level. This approach met with two obstacles. First, very few scholarship students have test scores across three time points that can be compared. This is largely due to missing test scores in the 2014-2015 academic year resulting from failures to report scores or students being in grades for which reporting was not required (e.g., K-2). Additionally, an individual student may not have taken the same standardized test each year (due to schools changing tests and students changing schools, especially from 8th grade into high school). A greater proportion of students can be compared over a two-year period and this approach was taken in some of the analyses below. A second issue is that changes can only be observed as state-wide gains or losses over academic years with the ALSDE data, which may obscure the actual amount of change occurring for individual students. For example, if proficiency rates remain constant from year to year, it is not clear whether that is due to there being no changes in individual student scores or if instead that the percentage of students who gained in proficiency was off-set by a similar percentage who dropped in proficiency. With no other viable alternative, for the ALSDE data, only change in proficiency rates between consecutive two-year periods can be observed and this was used in this report.

With these limitations in mind, three strategies were used to examine change over time.

- The first approach examined the relationship between the number of years a student had received a scholarship and their achievement test scores for the 2016-2017 academic year. This analysis includes the greatest number of scholarship students and test types, but it does not reveal the amount of change over time, only the direction of change.
- A second approach compared student performance across pairs of adjacent years for the three achievement tests in which there were sufficient numbers of students. These analyses examined how percentile scores in each of the subject areas changed on average for each student, taking into account their grade level, and can reveal if the amount of change was significant. For example, this analysis can reveal if the average 3rd grader who took the Iowa Test in the 2015-2016 academic year improved significantly in 4th grade for the 2016-2017 academic year. Although this approach includes a significant proportion of the scholarship students, it still excludes a number of students with missing test scores or who did not take the same test in consecutive years.
- The final approach focuses on changes in the ACT Aspire proficiency status in grades 3 - 8 because these scores can be directly compared to the results for Alabama public school students who also took the ACT Aspire. Conclusions from these comparisons, however, must be regarded with the caveat that scholarship students who took the ACT Aspire may not be representative of the scholarship students who took other tests. In addition, because 9th grade

students were not required to take the ACT Aspire, and because students did not generally take the same tests in 10th and 11th grade, high school students could not be included in this analysis.

Correlations between 2016-2017 Test Performance and Number of Years Receiving a Scholarship

Correlation analysis was used to infer a relationship between performance on the 2016-2017 achievement tests and the number of years receiving a scholarship. This analysis included all students who took one of the seven tests included in the report for Objective 1. Correlations were calculated between number of years a student had received a scholarship and their scores in math, reading, and English/language arts. A significant positive correlation would indicate that scholarship students' performance was higher the longer they were in the program, but a negative correlation would indicate that their achievement was lower the longer they were in the program. Non-significant correlations would suggest that there is no relationship between achievement test scores and the number of years a student had received a scholarship. Out of the 24 correlations calculated, only one was significant: math percentile scores on the PSAT/NMSQT were negatively correlated with the number of years receiving a scholarship (r = -0.25), suggesting that poorer performance was associated with being enrolled in the scholarship program over time. It should be noted that correlation does not mean causation. Given the data available, it cannot be determined whether performance on the PSAT/NMSQT is a result of receiving a scholarship or some other factor. More importantly, the overall majority of the correlation findings suggest that there is no relationship between performance on standardized tests and the number of years a student had received a scholarship.

Changes in Percentile Scores 2015-2016 vs 2016-2017

Comparisons were made between students' percentile scores for the 2015-2016 academic year and their scores for the 2016-2017 academic year for those who took the ACT Aspire, Iowa Test, Stanford-2002 norms, and Stanford-2007 norms. These were the only achievement tests with sufficient numbers of students to make a comparison across these years. As stated previously, there was a large amount of missing data in 2014-2015. A paired *t*-test was used to compare the percentile scores in English/language arts, math, and reading.

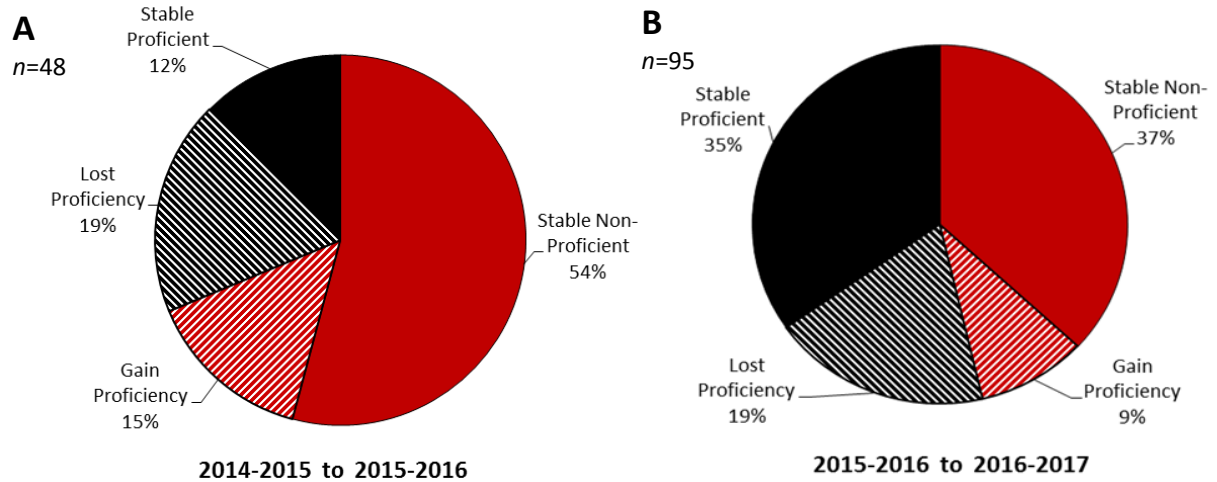
Overall, the results indicated no significant differences in percentile scores between 2015-2016 and 2016-2017, except in math and English on the Stanford-2002. In this case there was a significant decline in performance over time. It should be repeated that only a small proportion of the scholarship students are represented in the Stanford-2002 analyses and the results should not be generalized to all scholarship recipients. Across all four achievement tests, only two of the 12 comparisons made revealed a significant decrease in scores, so the broader pattern suggests that there was no significant change in scores from one year to the next.

Changes in ACT Aspire Proficiency

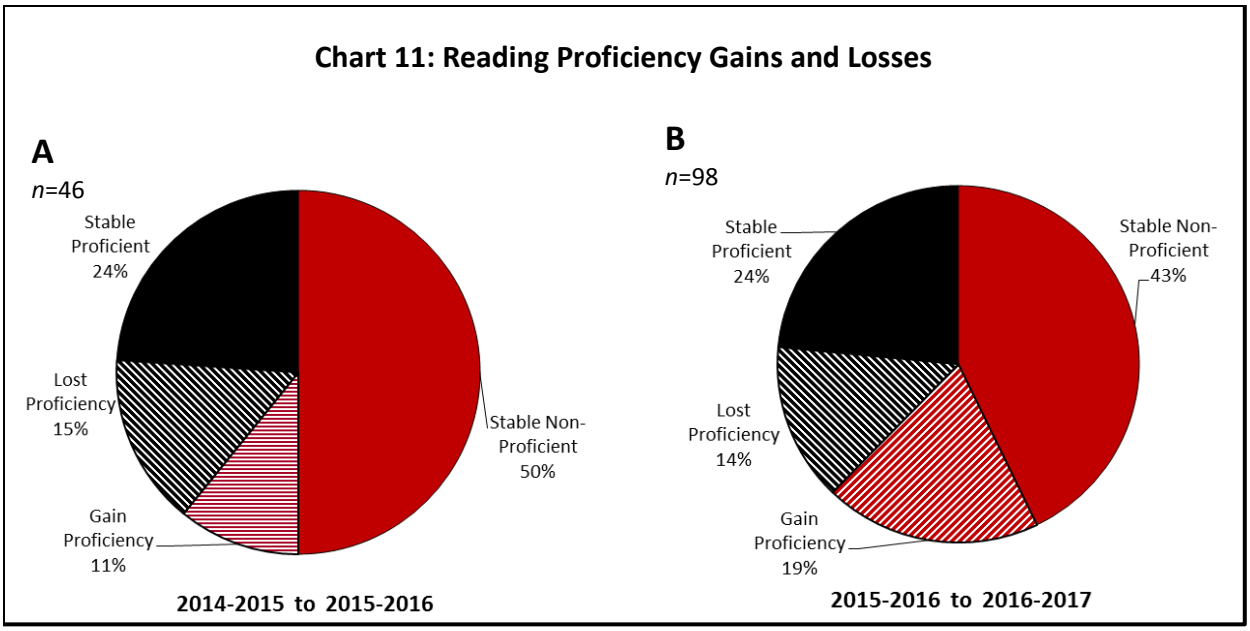
Change in proficiency status was examined across two sets of adjacent academic years 2014-2015 to 2015-2016 and 2015-2016 to 2016-2017. Within these time periods each student was placed into one of four groups for each subject area: Stable Proficient, Stable Non-Proficient, Gained Proficiency, and Lost Proficiency. Charts 10 and 11 below graphically represent these groups for math and reading. In each of these charts, the solid red and black pie pieces represent the percentage of students whose proficiency status did not change, with red indicating that they remained non-proficient and black indicating that they remained proficient. The striped pie slices indicate students whose proficiency status changed, with red stripes indicating a move from non-proficient to proficient (Gained Proficiency) and black stripes indicating a change from proficient to non-proficient (Lost Proficiency).

Considering math first, Charts 10A and 10B indicate that the majority of students did not change proficiency status (combined solid red and black), regardless of where they were classified in the previous year (66% in Chart 10A and 72% in Chart 10B). For the 2014-2015 academic year (Chart 10A) a larger number of students declined in their performance (19% - black stripes) than improved (15% - red stripes) into the next academic year. Chart 10B shows a similar trend for changes from 2015-2016 to 2016-2017 in that fewer students improved (9% - red stripes) than declined (19% - black stripes).

Chart 10: Math Proficiency Gains and Losses



For reading, a very similar pattern is observed. Charts 11A and 11B indicate that the majority of students did not change their proficiency status over time. In Chart 11A the percentage who lost proficiency (15% - black stripes) was greater than those who gained proficiency (11% - red stripes). The opposite pattern is seen in changes from 2015-2016 to 2016-2017, in that more students improved (19% red stripes) than declined (14%, black stripes).



Considering the patterns across both subject areas, the majority of students who were categorized as non-proficient one year, remained non-proficient the following year. However, a positive sign is that for the most recent comparison years, more students gained proficiency in reading than lost it. It is important to recall that this is only a subsample of the larger population of students who took the ACT Aspire in 2016-2017. The proficiency rates for this larger group including all grade levels was 46% for math and 38% for reading. Thus, the improvements suggested in these charts must be placed in the context that the majority of students in 2016-2017 were not rated as proficient in either subject area. As more data are collected among the scholarship students in the coming years a clearer picture will emerge as to the likelihood of improvement over time.

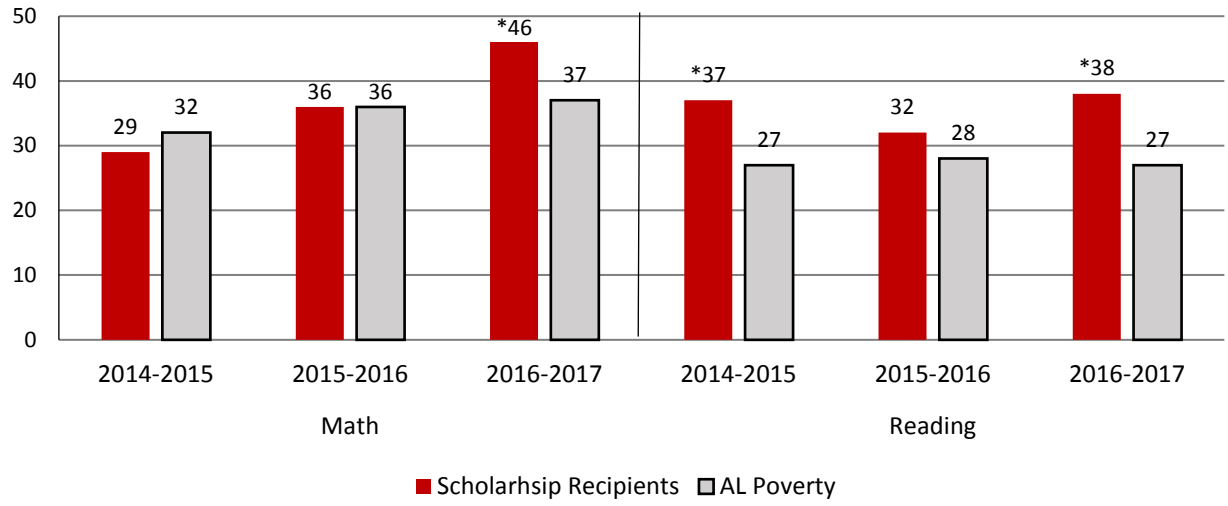
As noted earlier, ALSDE does not provide data that permits a similar analysis for students attending public schools. However, year-to-year state-wide changes in proficiency rates can provide an assessment of overall change in public school children's performance. The Alabama poverty group is most similar to the scholarship students, and this group is the focus of this summary. Chart 12 shows the average percentage of students (grades 3 - 8) meeting proficiency benchmarks in math and reading for the poverty group. Comparable data are included for scholarship recipients for each of the three academic years that were included in this report. It is important to note that the data previously presented in Charts 10 and 11 (*n*'s range from 46 to 98) only represent scholarship recipients who had ACT Aspire test scores for *two consecutive years*, and this excludes many students. Chart 12 compiles the scholarship students' proficiency rates for each academic year, and represents all scholarship students in grades 3 - 8 with ACT Aspire scores each year (*n*'s range from 122 to 286).

For the poverty group there was a four percentage point increase in math proficient students over the first two years presented, but all other proficiency rates were within 1 percentage point each year. This suggests that generally the achievement rate of the State poverty sample was stable over time. This analysis should not be interpreted as suggesting that individual children do not move from non-proficient to proficient or vice versa over academic years. Instead Chart 12 indicates that

if individual children in the poverty sample move from non-proficient to proficient, at the state level this improvement is offset by a similar number of children changing in the opposite direction.

In comparison, for the scholarship students, the percentage of students who were proficient in math for the 2016-2017 academic year was significantly higher than the percentages in 2014-2015 and 2015-2016, but there was no significant change from 2014-2015 to 2015-2016. There were also no significant changes in proficiency rates for reading. Overall, compared to Alabama poverty group, scholarship recipients' math scores seem to be going up during the years being compared, but reading scores do not show a consistent trend of improvement. A comparison of proficiency rates between the Alabama poverty and scholarship recipient students for each year revealed that scholarship recipients had a statistically significant higher proficiency rate than Alabama poverty students for both math and reading for the 2016-2017 academic year, consistent with findings reported for Objective 2. Scholarship recipients also had higher reading proficiency rates for the 2014-2015 academic year. Although these are positive finding for the AAA program, these comparisons need to be put in context of the low levels of proficiency for both groups of students across all years.

Chart 12: Average Percent Proficient Over 3 Years for Alabama Poverty and Scholarship Recipients



* Scholarship recipients had a significancy higher rate of proficiency than the Alabama poverty group.

Objective 3 Conclusions

Overall, the results indicate that over time participating in the scholarship program does not, on average, yield a significant improvement on standardized tests scores. Generally, the number of years that a student participated in the scholarship program was not correlated with achievement performance. Results indicated that for the majority of scholarship recipients there was no gain or loss in percentile scores on the ACT Aspire, Iowa Test of Basic Skills, and Stanford. On the ACT Aspire students were more likely to remain in a non-proficient category than to improve. While proficiency rates are typically well below 50%, there is some evidence that in the most recent year of this report, that scholarship students collectively improved in proficiency rates in math on the ACT Aspire. In 2016-2017, they also had a higher rate of proficiency in math and reading on the ACT Aspire compared to the Alabama poverty students. However, this group of scholarship recipients is only a small percentage of the larger group of participants in the AAA. The more consistent pattern across all scholarship students is a lack of change over time in achievement. This general lack of change over time follows the same pattern seen in public school students in Alabama and is likely not attributable solely to the scholarship program. It is important to be reminded that the analyses conducted for this objective assess patterns of change for the group as a whole. There are no doubt children who improve in their achievement over time in both the scholarship program and the Alabama public school system. However, the findings of this report suggest that the proportion of children who improve maybe offset by a comparable proportion who decline. The objective of this report is to summarize the patterns for the group as a whole, and these patterns do not indicate that as a group students' achievement test scores significantly improve or decline as they move through school.

Summary for Objective 3: Changes in Achievement across Time

- **On average, over time, participating in the scholarship program was not associated with significant improvement on standardized tests scores.**
- **The lack of change over time followed the same pattern seen in public school students in Alabama and is likely not attributable to participation in the scholarship program.**

✓ The **number of years** that a student participated in the scholarship program was ***not* correlated with higher** achievement test scores.

✓ For the majority of scholarship recipients there was **no gain or loss in percentile scores** on the **ACT Aspire, Iowa Test of Basic Skills,** and **Stanford 2007 norms.** Scores on the **Stanford 2002** declined for math and English.

✓ On the **ACT Aspire** students were more likely to **remain in a non-proficient** category than to improve. However proficiency rates for 2016-2017 were higher than those of Alabama poverty students.

General Conclusion

The purpose of the evaluation is to assess how the scholarship program enacted through the AAA affects the academic achievement of students in the program. Throughout the report many concerns have been voiced about the reliability and validity of the findings due to unknown factors associated with missing achievement tests and due to issues related to subsamples included in specific comparisons, such as whether a subsample of students accurately represented the larger group of scholarship students. Within these limitations, the report made use of the available information to describe how well the scholarship recipients in the 2016-2017 academic year performed. The evaluation addressed three objectives to reach this goal:

- The first objective described the achievement test results of the scholarship recipients and revealed that generally these students did not perform as well as other students in the U.S. Other indicators, such as the NAEP assessments, are consistent with these results, finding that students in the state of Alabama do not perform as well as students elsewhere in the country.
- When compared to Alabama public school students on ACT Aspire and ACT scores in Objective 2, there was no consistent pattern indicating that one group performed better or worse across grade levels. Only a small percentage of students took the ACT Aspire or the ACT, which hampers the ability of this report to draw definitive conclusions.
- Finally, the evaluation assessed if scholarship recipients' achievement scores improved, declined or remained the same over time. Similar to their public school counterparts, findings suggested that, on average, scores showed little improvement over time.

Limitations

The types of descriptive analysis found in this report will always have shortcomings in that they do not control for the litany of possible confounding differences among students and the schools they attend. This includes potential differences in test or grade samples, many of which have already been discussed, such as different compositions of race, household income, or number of years receiving a scholarship. The inability to control for these factors is largely due to limitations in the information/data that is available.

Creating an accurate model of the effects of the scholarship program would require statewide, student-level testing results that are linked to a student's demographic information. The most meaningful comparison between scholarship recipients and public school students would compare scholarship students' performance to the performance of students in the public school for which they were zoned, rather than aggregating across all schools in the state. Unfortunately, the state does not receive the student-level testing results needed for this approach. Gathering such information for each individual student would be time intensive and costly.

Drawing conclusions regarding the academic achievement of scholarship recipients relative to students attending public schools in no small part depends on the number of schools with scholarship recipients that use tests that are utilized by ALSDE in the future. ALSDE discontinued the use of the ACT Aspire for the 2017-2018 academic year, and there may be an additional test change in the 2018-2019 academic year. This will further constrain comparisons between

scholarship recipients and students attending Alabama public schools and to make comparisons over time.

Finally, it is important to reiterate that the use of proficiency scores to discern differences in student performance may not be sensitive to meaningful changes in performance. Proficiency scores reduce test scores to four-point scales, and change in performance is only registered when students transition from one group to the next. From a policy perspective, a considerably smaller change in scores could be considered significant. Additionally, students who are closest to the cutoff scores are more likely to change proficiency groups, entailing that a relatively small number of students can have a disproportionate impact on the results. A better understanding of student academic gains could be achieved by either using student-level testing results, or by knowing the means and other statistical information for test scores across demographic groups**.**

Comparison to Previous Reports

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 the U.S. in academic achievement. Thus, even the most effective interventions might be challenged to show dramatic improvements in a short period of time. A common theme across the three annual reports to date is that scholarship students, similar to their public school counterparts, often lag behind their peers in the country. This is the first year that change in achievement performance over time was assessed, and results indicated that, on average, achievement test scores are fairly stable among scholarship students similar the state as a whole. In the upcoming years, analyses will evaluate if these trends continue.

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 50th 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.

Appendix

Table 1A: Mean Stanford Achievement Test Scores for Grades 3 - 8 (2007 Norms)							
Grade	Group (N)	Math		Reading		Language	
		Scale Score	Percentile	Scale Score	Percentile	Scale Score	Percentile
3	All (49-52)	610	41	616	32	613	33
	AA/Black (34-37)	609	39	614	30	608	29
	Female (29-30)	612	43	620	36	621	41
	Male (<25)	*	*	*	*	*	*
4	All (46-47)	608	29	624	30	622	36
	AA/Black (30-31)	605	27	621	29	623	37
	Female (< 25)	*	*	*	*	*	*
	Male (27-28)	604	26	620	28	612	28
5	All (36-38)	631	27	635	24	627	25
	AA/Black (25-27)	624	20	626	17	617	18
	Female (< 25)	*	*	*	*	*	*
	Male (< 25)	*	*	*	*	*	*
6	All (45)	654	37	654	31	642	30
	AA/Black (31)	644	28	646	23	637	25
	Female (20-22)	*	*	*	*	*	*
	Male (≤ 25)	*	*	*	*	641	29
7	All (56-58)	656	28	655	23	647	28
	AA/Black (39-41)	649	21	645	16	641	23
	Female (26-27)	652	24	646	16	648	29
	Male (30-32)	660	32	663	31	645	27
8	All (67-69)	667	32	667	27	650	27
	AA/Black (52-54)	662	26	662	22	645	23
	Female (35-36)	673	37	676	36	665	38
	Male (32)	660	24	658	19	633	15

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

Table 2A: Mean Iowa Test Scores for Grades 3 - 8				
Grade	Group (N)	Math Percentile	Reading Percentile	English Percentile
3	All (194-196)	45	43	47
	Black (98-99)	33	35	39
	Hispanic (36)	62	42	49
	White (45-46)	55	58	61
	Male (90-93)	45	42	44
	Female (103-104)	44	44	49
4	All (166-171)	38	41	46
	Black (88-91)	28	33	38
	Hispanic (33)	45	46	50
	White (30-32)	51	55	62
	Male (94-98)	36	39	42
	Female (72-73)	40	44	52
5	All (130-136)	38	40	44
	Black (67-73)	27	33	35
	Hispanic (30)	51	47	48
	White (< 25)	*	*	*
	Male (56-60)	37	38	39
	Female (74-76)	38	42	48
6	All (149-151)	31	37	41
	Black (86-87)	24	30	32
	Hispanic (31)	32	35	40
	White (< 25)	*	*	*
	Male (86)	32	38	40
	Female (63-65)	29	35	41
7	All (157-162)	32	37	39
	Black (96-101)	27	33	33
	Hispanic (29)	39	33	37
	White (< 25)	*	*	*
	Male (69-72)	33	36	35
	Female (88-90)	32	37	42
8	All (174-181)	36	43	43
	Black (109-111)	28	37	36
	Hispanic (21-23)	*	*	*
	White (27-30)	47	57	59
	Male (94-95)	37	42	42
	Female (80-86)	35	43	44
* Indicates an insufficient number of students in the group (< 25) for reporting.				

Table 3A: ACT Aspire Proficiency Groups for Grades 3 – 8 and 10

Grades	Group (N)	Math Proficiency Groups				Reading Proficiency Groups				English Proficiency Groups			
		1	2	3	4	1	2	3	4	1	2	3	4
3	All (54)	9	33	46	11	39	30	21	11	4	21	37	39
	Black (40)	10	33	45	13	38	30	23	10	3	13	45	40
	Female (31)	10	36	42	13	39	29	16	16	3	13	45	39
4	All (39-40)	8	35	50	8	15	48	25	13	5	13	39	44
	Black (27)	11	33	44	11	19	52	19	11	7	19	33	41
	Female (*)	*	*	*	*	*	*	*	*	*	*	*	*
5	All (37-38)	13	50	29	8	26	58	16	0	0	32	32	35
	Black (*)	*	*	*	*	*	*	*	*	*	*	*	*
	Female (26-27)	15	48	26	11	22	59	19	0	0	31	27	42
6	All (37-40)	3	33	33	30	23	38	30	10	3	16	35	46
	Black (*)	*	*	*	*	*	*	*	*	*	*	*	*
	Female (28)	4	29	39	29	14	39	39	7	0	12	42	46
7	All (54-57)	21	44	25	11	18	44	38	0	2	13	41	45
	Black (32)	13	53	16	19	16	41	44	0	0	16	47	37
	Female (35)	26	46	26	3	18	46	36	0	3	12	36	49
8	All (54-57)	32	40	20	7	16	28	46	11	0	8	35	58
	Black (27-28)	39	39	14	7	25	29	39	7	4	4	48	44
	Female (33)	37	36	15	12	18	36	30	15	3	6	39	52
10	All (39-43)	71	24	5	0	40	30	26	5	14	33	33	21
	Black (26-29)	79	21	0	0	48	17	31	3	17	38	27	17
	Female ($\leq$ 25)	*	*	*	*	32	32	32	4	4	28	40	28

* Indicates an insufficient number of students in the group (< 25) for reporting.
Proficiency Groups: 1--*In need of support*, 2--*Close*, 3--*Ready*, and 4—*Exceeding*

Table 4A: Math and Reading: Comparison of the Percentage of Scholarship and Alabama Public School Students in Proficiency Groups for the ACT Aspire

	Math Proficiency				Reading Proficiency			
Grade/ Group	1 Needs Support	2 Close	3 Ready	4 Exceeds	1 Needs Support	2 Close	3 Ready	4 Exceeds
3rd Grade (54) Scholarship	9	33	46	11	39	30	21	11
AL Poverty	19	31	36	13	47	25	17	11
AL State	15	26	38	21	39	24	21	17
4th Grade (39-40) Scholarship	8	35	50	8	15	48	25	13
AL Poverty	11	47	34	8	37	33	21	9
AL State	9	39	37	15	29	31	25	15
5th Grade (37-38) Scholarship	13	50	29	8	26	58	16	0
AL Poverty	13	52	28	7	43	30	17	10
AL State	10	44	32	13	35	29	21	15
6th Grade (37-40) Scholarship	3	33	33	30	23	38	30	10
AL Poverty	14	42	31	13	41	27	20	13
AL State	10	35	32	23	32	24	23	21
7th Grade (54-57) Scholarship	21	44	25	11	18	44	38	0
AL Poverty	37	37	18	8	42	33	22	4
AL State	29	33	22	16	33	31	28	8
8th Grade (54-57) Scholarship	32	40	20	7	16	28	46	11
AL Poverty	48	30	14	8	33	30	27	9
AL State	36	29	18	17	26	28	31	15
10th Grade (39-43) Scholarship	71	24	5	0	40	30	26	5
AL Poverty	76	14	7	3	54	23	19	4
AL State	63	17	12	8	44	24	25	8

Percentages within a category may sum to a total slightly different from 100 due to rounding.