Evaluation of the Alabama Accountability Act: Academic Achievement Test Outcomes of Scholarship Recipients for 2022-2023

The Institute for Social Science Research
The University of Alabama

Erika Steele, Ph.D.
Joan M. Barth, Ph.D.

September 1, 2024

Executive Summary

This report fulfills the Alabama Accountability Act (AAA) evaluation requirements by examining the academic achievement of scholarship recipients through the 2022-2023 academic year.

Report 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 (SGOs) provided demographic information and achievement test scores for scholarship recipients. Achievement test score information for Alabama public school students was retrieved from the Alabama State Department of Education (ALSDE) website, the Public Affairs Research Council of Alabama, and the ACT Inc.

- The SGOs provided information on 2806 scholarship recipients for 2022-2023.
- Recipients attended 118 schools in 46 counties in Alabama.
- 1925 recipients were in grades 2 through 8, 10, and 11 and required to submit test scores.

Methodological Limitations

- Accurate assessment of the scholarship recipients' academic achievement and the comparisons to Alabama public school students are limited due to the lack of a uniform standardized test among schools.
  - Norm-referenced tests and criterion-referenced tests are based on different standards and cannot be directly compared.
  - Schools using the same test often reported scores based on different national norms and these cannot be combined.
  - Some achievement tests were used by only one school or included only a small number of students, making analyses unreliable.
  - Since 2020-2021, Alabama public school children in grades 2 through 8 have taken the Alabama Comprehensive Assessment Program (ACAP), and thus there is limited longitudinal data on their performance. Because only a few AAA scholarship students took this test, no direct comparisons could be made for these grades.
- Inconsistencies in test score reporting from schools and missing test data limited the number of students who could be included in the evaluation sample.

2022-2023 Sample Characteristics

After accounting for the issues noted, the evaluation was based on 1429 scholarship recipients. This represented 74% of the scholarship recipients in the grades for which testing was required. Student demographic characteristics included the following:

- 51% were female.
- Number of years receiving a scholarship:
  - 29% were first-time scholarship recipients.
  - 45% were in their 5th year or more as a scholarship recipient.
- 93% were eligible for free/reduced lunch subsidies.
- 62% were Black, 18% were White, and 17% were Hispanic.

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

- On norm-referenced tests, scholarship students as a group did not perform as well as students in the U.S. taking the same test and did not meet grade level academic standards.
  - Typically, mean percentile scores across tests were significantly below the 50th percentile.
  - There were some exceptions in which mean scores were not below the 50th percentile for small numbers of students on specific tests. Variability among the grades, subject areas, and demographic groups in which these scores occurred revealed no reliable pattern.
  - Achievement indicators on norm-referenced tests revealed that most students did not meet academic standards for their grade level and that performance declined from 3rd to 8th grade.
  - Variability in achievement outcomes on different tests is likely due to differences among schools that choose specific tests, such as school demographics and pedagogical approaches.
- Most criterion-referenced test included in this report were taken by students in grades 10 and 11.
  - Across all tests and grade levels the majority of students failed to meet Math benchmarks.
  - Findings for English and Reading were mixed and depended on the test and grade level.
    - PSAT/NMSQT Reading-Writing assessment: the majority of 10th and 11th graders met proficiency standards.
    - PreACT 10th grade: the majority met proficiency standards in English, but not Reading.
    - ACT 11th grade: the majority failed to make benchmarks scores in Reading and English.
  - More reliable results are likely represented in the ACT and the Pre-ACT tests because they include a greater percentage of scholarship recipients compared to the PSAT.

Objective 2: Compare the Learning Achievement of Scholarship Recipients to Students Attending Public Schools

- The majority of economically disadvantaged public school students failed to meet test specific grade level benchmarks for Language Arts and Math.
- Due to the lack of comparative data for grades 2 through 8, strong conclusions cannot be made for the relative performance of the scholarship recipients and public school students.
- For 11th graders taking the ACT, proficiency rates were higher for scholarship recipients in English compared to economically disadvantaged public school students, but there were no differences between these groups in proficiency rates for Math.

Objective 3: Assess Changes in Achievement Across Time

- On average, the number of years a student participated in the scholarship program was not strongly correlated with significant improvement on standardized test scores.
- ACT scores have not significantly improved since 2016 for scholarship students**,** with the exception of Reading. However, there has been no significant and sustained improvements in Reading since 2018.
- Starting in 2021, the mean ACT Reading scores for the scholarship recipients were significantly higher than the economically disadvantaged public school students. In 2022 and 2023, the English and Math scores were also significantly higher for the scholarship recipients.

Table of Contents

List of Abbreviations

AAA	Alabama Accountability Act
ACAP	Alabama Comprehensive Assessment Program
AL	Alabama
ALSDE	Alabama State Department of Education
COVID-19	Coronavirus disease 2019
Econ. Dis/Disadv	Economically Disadvantaged
ELA	English Language Arts
EBRW	Evidenced Based Reading-Writing
FERPA	Federal Education Rights and Privacy Act
ISSR	Institute for Social Science Research
K-12	Kindergarten through 12th grade
N	Number of people in a group
n	Number of people in a subgroup
NA	Not applicable
NAEP	National Assessment of Educational Progress
PARCA	Public Affairs Research Council of Alabama
PDF	Portable Document Format
PSAT/NMSQT	The Preliminary SAT/National Merit Scholarship Qualifying Test
S	Spring norms, for example Iowa S2017 means Iowa Spring 2017 norms
SAT-10	Stanford Achievement Test-10
STAR	Standardized Testing and Reporting Assessment
SGO	Scholarship Granting Organization
r	Correlation coefficient

List of Charts and Tables

Evaluation of the Alabama Accountability Act: Academic Achievement Test Outcomes of Scholarship Recipients through 2022-2023

Introduction

The Alabama Accountability Act (AAA), passed by the legislature in 2013, established a statewide scholarship program for low-income students to attend public or private schools instead of the public school for which they are zoned. This report fulfills the AAA mandated evaluation of the academic outcomes of students receiving scholarships as set forth in the AAA legislation. It follows a series of biannual reports starting in 2016 authored by the Institute for Social Science Research (ISSR) at the University of Alabama that described the achievement test results for scholarship recipients. The 2024 report examines these same issues for the 2022-2023 academic year.

This report first provides an overview of the pertinent AAA legislation. The methodology is described next, which includes a description of the 2022-2023 sample and the achievement tests that are part of this report. The findings are organized around three objectives set forth in the legislation: a) describe the academic achievement of students receiving tuition scholarships in the 2022-2023 academic year, b) compare scholarship recipients' performance to public school children, and c) examine changes in achievement over time. The conclusion of the report summarizes the overall impact of the AAA scholarship program on student academic achievement.

Overview of AAA

The AAA scholarship program for low income students is funded by a tax credit program. The scholarship awards are managed by Scholarship Granting Organizations (SGOs), which must comply with standards set by the AAA. All students receiving scholarships must meet family income eligibility requirements. First consideration is given to students who are zoned to attend a public "priority school" (formerly referred to as a failing public school) as designated by the Alabama State Department of Education (ALSDE). However, students meeting AAA income requirements who attend public schools that are not priority 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 to a student depending on the grade level (elementary, middle, or high school). Additional amendments in 2023 impacted income eligibility and testing requirements, but these were not in effect for the current reporting period. The Alabama State Department of Revenue oversees the implementation of the 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 2022-2023 academic year.

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 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.

Following these requirements, each evaluation report since 2016 has had the same three objectives: a) describe the academic achievement of students in the scholarship program for the most recent academic year for which data are available (2022-2023 for the current report), b) make comparisons between the level of achievement of the scholarship recipients and comparable students attending public schools, and c) measure the achievement gains of students in the scholarship program over time.

Alabama State-Mandated Testing in Public Schools for the 2022-2023 Academic Year

The Alabama State Department of Education assesses children in grades 2 through 8 using the Alabama Comprehensive Assessment Program (ACAP). ACAP is an online assessment designed to provide state stakeholders with information regarding student progress toward mastery of the Alabama Course of Study Standards. Publicly available ACAP reports for students in grades 2 through 8 include English Language Arts and Math. Additionally, 2nd and 3rd grade Reading scores are available. Alabama tenth graders took the PreACT, and eleventh graders were required to take the ACT college entrance exam. Tests were typically administered during the spring semester in March and April.

Impact of COVID-19

The COVID-19 global pandemic had an enormous impact on education in Alabama and the entire U.S. during the 2019-2020 school year. The State of Alabama closed all public schools for in-person instruction in March 2020. On March 27, 2020, the U.S. Department of Education approved the state's request to waive federally required student assessments and other measures of student achievement for students in grades K-12. Thus, standardized testing, including college achievement and entrance exams, that typically occurs in the latter half of the spring term was cancelled. The majority of private schools attended by scholarship recipients were also closed during this time and consequently did not test students. The lack of test data for the 2019-2020 academic year impacts Objective 3 of this report, which examines the change in scholarship students' academic achievement over time. Strategies developed for the 2022 report to account for the missing test scores continue to be utilized in this report.

Method

The methodology for the 2024 report follows that of previous years, and similarly, the conclusions that can be drawn from this report are limited in several ways by the nature of the testing data that are reported to the evaluation team. These are briefly discussed as they remain largely unchanged from previous reports.

The lack of a uniform achievement test among schools profoundly affects the conclusions that can be made about student achievement outcomes and the types of comparisons that can be made to students attending public schools. Schools provided scores from 23 standardized tests. Comparisons across tests are invalid because tests vary in their content and are designed for unique purposes. *Norm-referenced tests*, such as the Iowa Assessment and the Stanford Achievement Test, and *criterion-referenced* tests, such as the ACT college entrance exam, are based on different standards and cannot be directly compared.

Additionally, some tests were used by only one school or taken only by a small number of students. Small numbers for some grade levels and demographic groups also make comparisons potentially unreliable. Guidance from ACT Inc. recommends a sample of at least 25 students, and this standard was adopted in this report (Sawyer, 1987).

An additional issue is that even when the same test is used across schools, scores could be based on different norm years. For example, one school may report test scores for the Stanford Achievement Test based on 2018 norms, while another school may report scores based on 2002 norms. Test results are not comparable because the older norms are not based on the Common Core, the current national standards for children in grades K-12. When outdated norms are used, scores do not provide an accurate accounting of students' achievement. The 2024 report only includes test score data based on the most recent norms available so that a more accurate assessment of scholarship students' academic performance can be given.

Every year the evaluation team communicates with the SGOs about the specific test scores that should be reported in the test reports, including the subject areas (Reading, Language/English, and Math) and types of scores (national percentiles and scale scores). These expectations are communicated to the schools. School adherence to these guidelines has improved over time, but missing data continues to compromise the integrity of the findings. With these challenges noted, the remainder of the report describes outcomes for the 2022-2023 academic year.

Data Sources

The following data sources were used to evaluate the academic achievement of the 2022-2023 scholarship recipients:

- Demographic reports from six SGOs: Alabama Opportunity Scholarship Fund, C2 Opportunity Scholarship Fund, Children's Tuition Fund, Scholarships for Kids, Rocket City Scholarship Granting Organization, and Renaissance Scholarship Fund.
- Test reports collected by the SGOs from participating schools and shared with ISSR. Test scores were received as PDFs.
- 2022-2023 Alabama Comprehensive Assessment Program (ACAP) results available from the ALSDE website, Alabama Achieves.
- Eleventh (11^{th}) grade ACT scores for public school students in Alabama retrieved from the Public Affairs Research Council of Alabama (PARCA) report available on their website.

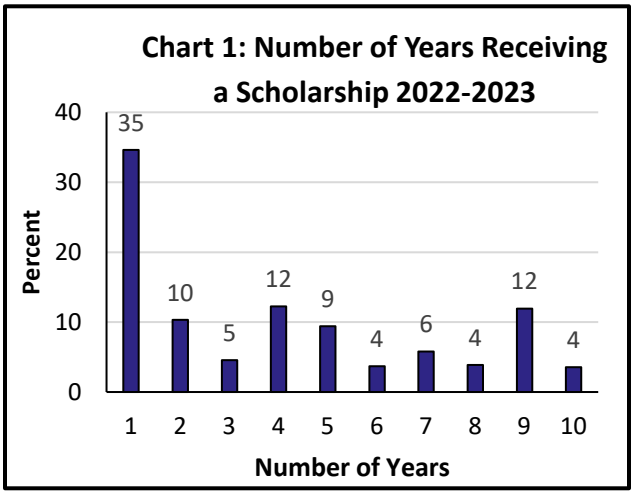
Statistical Analyses

Statistical comparisons were conducted throughout the report to aid in drawing conclusions. These statistical tests consider 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 $\leq .05$ was used as the criterion to determine significance.

- T-tests were used to compare mean scholarship student test scores to established benchmarks, to compare genders, or to compare racial/ethnic groups of scholarship students.
- Analysis of Variance (ANOVA) was used to compare the mean scores of multiple groups, such as racial groups, or multiple years.
- Chi-Square analyses were used to compare demographic groups on the percentages of students meeting a benchmark score.
- Z-tests were used to compare the percentages of scholarship students meeting benchmarks to comparable indicators for public school students.
- Correlations (*r*) assessed the relation between achievement test scores and the number of years of participation in the AAA scholarship program.

2022-2023 Sample

The SGOs provided information on students who had received scholarships during the 2022-2023 academic year. Based on the information provided in the SGO reports, 2806 students (49% female) in kindergarten through 12th grade received scholarships during the 2022-2023 academic year. Chart 1 graphically illustrates the number of years participants had been in the AAA scholarship program. Thirty-five percent (35%) of students were first-time scholarship recipients, 27% had received a scholarship for two to four years, and 39% had received a scholarship for five or more years. Nearly all students (94%) were eligible for free/reduced lunch (also referred to as the poverty rate). The racial/ethnic composition of the scholarship recipients was 63% Black, 19% White, 14% Hispanic, and 4% were another race or no information was provided. Students attended 118 different schools and resided in 46 unique counties in the State of Alabama.



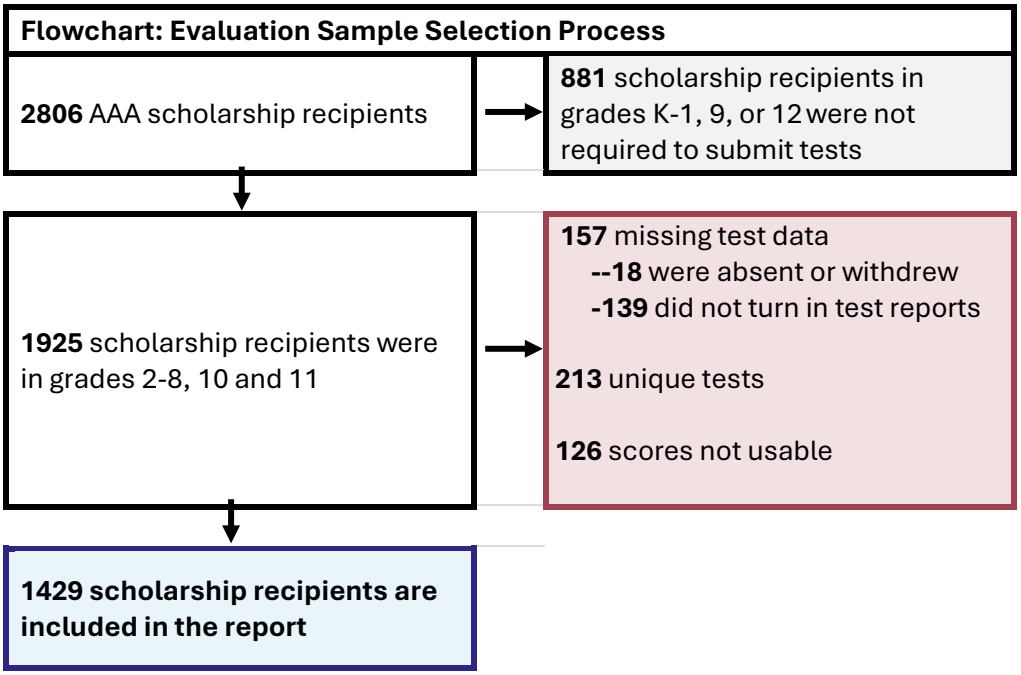
Achievement Test Data Included in the Report

Not all of the 2806 students who received scholarships had test scores included in this report. The flowchart below summarizes the evaluation sample selection process. Students in kindergarten, 1st, 9th, or 12th grades were not required to be tested in Math and Language Arts in the State of Alabama and made up 31% (*n* = 881) of scholarship recipients. Nearly all of the 1925 students (92%) who were in grades required to be tested had test scores submitted to ISSR by one of the SGOs.

Test scores were missing for 157 students (8% of those required to be tested) for a number of reasons: the student withdrew before testing (*n* = 10), the student was absent for testing (*n* = 8), the school did not submit scores to the SGO (*n* = 115), or the SGO provided no explanation (*n* = 24).

An additional 339 students' test results (18% of those required to be tested) could not be included in this report for multiple reasons. Several schools used unique tests (*n* = 213) with too few students taking them to provide a meaningful comparison. A single student taking a test may be identifiable, and this represents a violation of the Family Educational Rights and Privacy Act (FERPA). Other test reports (*n* = 126) were not usable for several reasons. Some (*n* = 14) of the test reports provided scores other than a national

percentile or scale score (e.g., percent correct, growth percentile rank). A significant number were not taken in the spring (9%; n = 30) or had out-of-date or other issues with the test norms (18%; n = 62). Tests taken in the fall do not reflect what the child has learned during the year, out-of-date norms do not align with the standards used by the State of Alabama or other States in the U.S. Finally, some schools submitted a test that was taken prior to the 2022-2023 school year (n = 11), six students had test reports that were illegible, and three students' test scores were not from an official report from a test company or agency. The final sample of scholarship recipients included 1429 students, which was 74% of the students for whom testing was required. ISSR will continue to work with the SGOs to ensure that all students who are in grades that are required to be tested in the State of Alabama take a valid standardized test.



Demographics for Scholarship Recipients Included in the Evaluation of Achievement Tests

Based on the information provided by the SGOs, the racial demographics of the 1429 scholarship recipients with usable test data did not significantly differ from the larger group of scholarship recipients. The racial composition was comparable: Black (62% vs. 63% in the full sample), White (18% vs. 19% in the full sample), Hispanic (17% vs. 14% in the full sample), and another race or none was designated (3% vs. 4% in the full sample). The gender composition was similar (51% vs. 49% in the full sample), as was the free/reduced lunch (poverty) rate (93% vs. 94% in the full sample). However, fewer first-year scholarship recipients were among the students required to be tested (26% vs. 35% in the full sample), and more students for whom testing is required had received scholarships for five or more years (45% vs. 38% in the full sample).

Description of Tests Included in the Report

Achievement tests are based on national educational standards like the National Assessment of Educational Progress (NAEP) and the Common Core State Standards. However, tests use different types of scores to evaluate how well students meet national or state standards.

- **Norm-referenced tests** (e.g., Iowa Assessments, Stanford Achievement Test) compare a student's performance to that of other students at the same grade level. These tests use national percentile scores to show how a student ranks compared to others. For example, a student at the 50th percentile is performing as well as or better than half of the students in the country at their grade level. Norm-referenced tests sometimes include proficiency groups to show if students are meeting grade-level standards.
- **Criterion-referenced tests** (e.g., The ACT, PSAT) focus on whether students meet specific achievement benchmarks, indicating if they are on track to reach long-term academic goals like college readiness. Unlike norm-referenced tests, where students are compared to each other, in theory, all students could meet the benchmarks on a criterion-referenced test. These tests use scale scores, which are based on the number of correct answers, to determine if students are meeting grade-level standards.

Because these two types of tests are designed differently, their results are presented separately.

Table 1 indicates the number of students and schools associated with each test used in the report. Seven standardized tests for scholarship recipients are included in this report: 1) Iowa Assessments, 2) Measures of Academic Progress (MAP) Growth, 3) PreACT, 4) ACT, 5) Preliminary Scholastic Aptitude Test-National Merit Scholarship Qualifying Test (PSAT/NMSQT), 6) Stanford Achievement Test 10 (SAT-10), and 7) Standardized Testing and Reporting Assessment (STAR). In addition, the report includes data from the Alabama Comprehensive Assessment Program (ACAP), which is taken by Alabama public school children in grades 2 through 8. MAP Growth, Iowa Assessments, SAT-10, and STAR are norm-referenced tests, and the PreACT, ACT, PSAT, and ACAP are criterion-referenced tests. Descriptions of each test follow.

Table 1: Tests Included in the Evaluation for Grades 2-8, 10, and 11

Test	Number of Students	Number of Schools*
Iowa Assessments	645	33
MAP Growth	388	24
SAT-10	76	14
STAR	31	8
PSAT/NMSQT	88	18
PreACT	99	18
ACT	104	22
Total	**1429**	
*Some schools used more than one test		

Norm-Referenced Tests

- The *Iowa Assessments* (previously Iowa Test of Basic Skills), currently administered by Riverside Insights, is a norm-referenced test. Test items, originally developed by the University of Iowa, align with Iowa Core State Educational Standards. The test has been validated at the national level, and it provides national percentile scores for Reading, Language, and Math. The scale scores can be used to track a student's progress over time. This report includes test results and interpretations based on national norms developed in 2017, the most recent at the time of testing.
- The *Measures of Academic Progress (MAP) Growth* is a computer-adaptive test developed by the Northwest Evaluation Association (NWEA). MAP Growth has features of both norm- and criterion-referenced tests. Scores are provided for Reading, Language, and Math. The National Achievement Percentile Ranks allow student performance to be compared to that of students in a norm group for the fall, winter, and spring terms. The **R**asch Un**IT** (RIT) Scores (100 to 350) help determine student proficiency levels based on cut scores set by individual states or default cut scores for U.S. states and international schools that have not set benchmarks. In 2022, NWEA conducted a study to create accurate benchmarks in Reading and Math for students in grades 3 through 8 for states that lacked established benchmarks (Tran et al., 2022). The new cut scores are based on student performance from states with set standards to establish accurate assessment of

grade level performance. These new benchmarks were used to assess the performance of AAA scholarship recipients, and National Achievement Percentile Ranks were used to compare student performance to that of other students in the nation.

- The *Stanford Achievement Test, 10th Edition (SAT-10)* is a norm-referenced test developed by Pearson Assessments. The SAT-10 compares a child's academic achievement relative to others in the nation based on a national percentile score. The SAT-10 provides national percentile scores in Language Arts, Reading, and Math for students in grades kindergarten through 12. Percentile scores of 24 or greater indicate average performance in Reading, Math, and Language Arts. Comparisons in the report are based on percentile scores using spring 2018 norms.
- The *Standardized Testing and Reporting Assessment (STAR)* developed by Renaissance Learning Inc. is a computer-adaptive test that has the qualities of both norm- and criterion-referenced tests. Like the other norm-referenced tests, national percentile rankings compare student performance to similar students in the nation. Grade level benchmarks are based on the 40th percentile for Math and Reading for students in grades 3 through 12.

Criterion-Referenced Tests

- The *Preliminary Scholastic Aptitude Test-National Merit Scholarship Qualifying Test (PSAT/NMSQT)* is used to prepare students to take the Scholastic Aptitude Test (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. The composite score is the sum of the Math and Evidence-Based Reading and Writing scores. Scores on these two subjects range from 160-760. Benchmarks are provided to assess students' college readiness.
- The *PreACT* and *PreACT 8/9* are used to prepare high school students to take the ACT college entrance exam. The scores can be used to predict how well a student might perform on the ACT college entrance exam. Reports include estimated ACT scores of 1-35 for PreACT and 1-30 for PreACT 8/9. Subscales are provided for Reading, English, and Math. Proficiency benchmarks are provided by ACT Inc. for grades 8 through 10 to assess college readiness. The PreACT is also used by the State of Alabama to assess the performance of 10th grade students.
- The *ACT* is a nationally normed college entrance exam, usually taken by 11th and 12th grade students to predict college readiness. Reports include an ACT composite score (1-36), which can be used to determine college readiness. Subscale scores are provided for Reading, English, and Math. ACT Inc. provides college readiness benchmarks and has set proficiency benchmarks for high school students. The State of Alabama also uses the ACT to assess the performance of 11th grade students and has its own proficiency benchmarks for student performance.
- The *Alabama Comprehensive Assessment Program (ACAP)* is a criterion-referenced assessment designed to measure grade level performance in English Language Arts and Math based on standards set by the Alabama Course of Study for students in grades 2 through 8. Students in grades 4, 6, and 8 are also assessed in science. Scale scores on the ACAP range from 200-850 and are used to determine grade level proficiency.

2022-2023 AAA Student Achievement Findings

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

Outcomes for each of the seven tests are described separately below. For each test, a brief description of the student demographics is provided with further details offered in Table 2 for those taking norm-referenced tests and Table 7 for those taking criterion-referenced tests. When possible ($n > 25$), test results are disaggregated by grade, race/ethnicity, and gender in tables. Due to rounding, sometimes percentages

in a table sum to a number greater or less than 100%. Statistical tests comparing racial/ethnic groups and genders are conducted when there are sufficient numbers of students in these groups. National percentile and scale scores are reported, as appropriate for the type of test. Additional information is provided (e.g., benchmark scores, test norms) to help interpret test results

The results are organized by test type—norm-referenced or criterion-referenced—since these assessments measure achievement differently. Summarized first are the norm-referenced tests (Iowa Assessments, MAP Growth, SAT-10, and STAR) and the summary of criterion-referenced tests (PSAT/NMSQT, PreACT, and ACT) follows. The AAA requires testing scores in Math and Language Arts. In some cases, English scores are reported instead of Language Arts, but the content is comparable. Additionally, Reading scores are included in this report because they were used in previous evaluations by ALSDE to assess achievement of public school children, and they have been included in the previous evaluation reports.

2022-2023 Norm-Referenced Test Results

Norm-referenced tests assess how students perform compared to others in the same grade in the United States. The 50th percentile is often used as an indicator of student performance. If scholarship recipients are performing at a similar level to their peers nationwide, their average scores should be close to the 50th percentile. Generally, meeting or exceeding this standard is seen as a positive outcome. However, percentile scores alone do not provide information on whether students have mastered the skills expected for their grade level.

Statistical comparisons were conducted for each test to see if the average scores in each subject were significantly above or below the 50th percentile. Some norm-referenced tests also offer benchmarks for grade-level proficiency, and this information is included when available. Table 2 provides a summary of the demographic characteristics of the students who took the norm-referenced tests.

Table 2: Student Demographics for Norm-Referenced Tests

Test	N	Race %				% Gender		% Poverty	% Years Receiving a Scholarship									
		Bk	Wh	His	Oth	F	M		1	2	3	4	5	6	7	8	9	10
Iowa	645	83	10	5	2	49	51	96	28	12	6	14	13	8	6	3	9	2
MAP	386	30	21	44	6	49	51	93	30	7	4	14	14	1	7	9	13	1
SAT-10	76	58	36	4	3	62	38	96	40	1	0	13	15	7	9	3	13	0
STAR	31	29	42	29	0	71	29	100	42	13	16	6	16	0	3	0	0	3

Bk = Black; Wh = White; His = Hispanic; Oth = Race Other includes multiracial, unspecified, and other races and ethnicities (e.g., Asian, Middle Eastern)
% poverty is the percent eligible for free/reduced lunch

Iowa Assessments

Results for the Iowa Assessment-Spring 2017 Norms were available for 645 students in grades 2 through 8, 10, and 11. The sample was predominantly Black (83%), and nearly all were free/reduced lunch eligible. Twenty-eight percent (28%) were first time scholarship recipients, and the number of years receiving a scholarship ranged from 2 to 10 for the rest of the students. (See Table 2.)

Table 3 shows the test results for each grade level, broken down by gender and race when there were enough students in each group. In grades 3 through 10, only Black students had sufficient numbers to report their scores. For gender, both males and females had sufficient numbers in grades 3 through 8, but only males had enough numbers in grade 10.

Most of the average percentile scores were significantly below the 50th percentile (marked with * in Table 3). However, there were a few exceptions where scores were not significantly different from the 50th

percentile: all scores for grade 2 (across all subjects and groups), grade 3 male Math scores, grade 5 male Language scores, and grade 8 female Language scores.

Table 3: Mean Iowa Assessment Percentile Scores and Achievement Levels for Grades 2-8 (Spring 2017 Norms)

Grade	Group (N)	Reading		Language	Math	
		Mean Percentile	% at Achievement Level	Mean Percentile	Mean Percentile	% at Achievement Level
2	All (29)	47	NA	43	61	NA
3	All (75)	39*	52	41*	41*	61
	Black (58)	39*	50	40*	41*	62
	Females (44)	39*	52	41*	35*	50
	Males (31)	40*	52	41*	49	77
4	All (87)	42*	56	43*	36*	48
	Black (70)	40*	51	40*	33*	41
	Females (49)	42*	53	43*	35*	53
	Males (38)	43*	61	43*	38*	42
5	All (67)	40*	52	43*	36*	51
	Black (51)	40*	49	42*	34*	49
	Females (40)	39*	48	41*	33*	48
	Males (27)	42*	59	45	40*	56
6	All (109-111)	33*	42	36*	24*	27
	Black (97-99)	31*	38	33*	22*	23
	Females (57-59)	34*	48	41*	22*	24
	Males (52)	32*	37	30*	26*	31
7	All (92-94)	35*	43	32*	24*	31
	Black (77-79)	33*	40	31*	22*	26
	Females (36)	36*	43	36*	24*	30
	Males (56-58)	34*	43	30*	24*	31
8	All (102-104)	36*	38	37*	25*	23
	Black (83-85)	32*	32	34*	22*	19
	Females (50-51)	43*	51	46	27*	24
	Males (52-53)	30*	24	28*	23*	23
10	All (47-50)	37*	42	39*	29*	24
	Black (42-44)	37*	43	40*	28*	23
	Males (34-35)	38*	46	39*	29*	29
11	All (26)	40*	42	39*	28*	15

*Mean score is significantly below the 50th percentile.
Mean scores without a * designation are not significantly different from the 50th percentile.
NA indicates that achievement levels were not available for 2nd grade.

Comparisons between scores for males and females at each grade level revealed three significant differences. Males scored higher than females in Math in 3rd grade, and females scored higher than males in Language in the 6th and 8th grades.

The Iowa Assessment also provides proficiency levels for Reading and Math to help interpret test results for grades 3 through 11. Table 3 indicates the percentage of students who reached a minimum level of achievement designated as "proficient" or higher. For Reading and Math, close to half or more than half of students met the proficiency benchmark in grades 3 through 5, but the remaining grade levels had 43% or fewer students reach proficiency in Reading and 31% or less reach the proficiency standards for Math.

Measures of Academic Progress (MAP) Growth

The MAP Growth test was administered to 388 students in grades 2 through 8, 10, and 11. The racial demographics of the MAP test-takers differed from the overall AAA scholarship population with Hispanic students having a relatively greater representation (44%) compared to Black (30%) and White students (21%). Thirty percent (30%) of the test takers received their first scholarship in the 2022-2023 school year, and the remaining students received a scholarship two to ten years. Nearly all (93%) of the students were eligible for free/reduced lunch (Table 2).

Table 4 presents the mean percentile scores for students in grades 3 through 8. There were less than 25 students in grades 10 and 11. Second-grade scores were also excluded because some test forms were missing specific scores in Reading, Language, or Math, reducing the number of second-grade students to less than 25. Scores were disaggregated by race and gender when the sample size was sufficient.

When examining the percentile scores for all students at each grade level, Reading scores were significantly below the 50th percentile, except for grades 3 and 6 where the mean scores were not significantly different from the 50th percentile. The Reading percentile scores for Hispanic students were significantly below the 50th percentile in grades 5, 7, and 8. Mean scores in Language were generally not significantly different from the 50th percentile, except for female 8th graders who scored above it and Hispanic fifth-grade students who scored below it. With respect to Math, mean scores for 3rd and 4th graders did not significantly differ from the 50th percentile. From fifth to eighth grade, mean Math scores were significantly below the 50th percentile, except for 7th grade boys, whose mean percentile score of 45 was not significantly different from the 50th percentile.

Table 4 indicates the percentage of students in grades 3 through 8 who met grade level achievement benchmarks for Reading and Math set by NWEA in 2022 (Tran e. al., 2022). For Reading, with the exception of girls in grade 3, less than 50% of students met grade-level benchmarks. For Math, 61% to 64% of students in third grade met the benchmark, but in the 4th grade the percentage of students meeting benchmarks dropped to 38% to 39%, and later grades had much lower percentages of students meeting the Math benchmarks (range 7% to 29%).

Table 4: Mean MAP Growth Percentile Scores and Percentage Meeting Grade Level Benchmarks for Grades 3-8

Grade	Group (N)	Reading		Language	Math	
		Mean Percentile	% Meeting Benchmark	Mean Percentile	Mean Percentile	% Meeting Benchmarks
3	All (51-67)	51	42	47	50	62
	Hispanic (26-31)	51	46	52	51	62
	Male (31-38)	48	34	48	50	61
	Female (20-28)	54	54	---	51	64
4	All (24-50)	44*	30	51	45	39

Table 4: Mean MAP Growth Percentile Scores and Percentage Meeting Grade Level Benchmarks for Grades 3-8

Grade	Group (N)	Reading		Language	Math	
		Mean Percentile	% Meeting Benchmark	Mean Percentile	Mean Percentile	% Meeting Benchmarks
	Female (25-26)	48	35	54	45	38
5	All (49-52)	42*	27	46	35*	13
	Hispanic (28)	37*	18	40*	31*	7
	Female (29-31)	44	29	48	36*	13
6	All (45-51)	48	43	50	36*	18
	Male (24-27)	47	44	---	33*	11
7	All (49-50)	43*	38	51	40*	26
	Hispanic (29)	40*	34	48	39*	24
	Male (27-28)	42	36	55	45	29
8	All (67)	44*	12	50	38*	16
	Hispanic (31)	40*	14	52	38*	14
	Male (29)	41*	10	41	40*	21
	Female (38)	46	13	57**	36*	13

* Mean score is significantly below the 50th percentile.
** Mean score is significantly above the 50th percentile.
--- Score not reported due to sample size < 25.
Mean scores without a * designation are not significantly different from the 50th percentile.

Stanford Achievement Test 10 (SAT-10)

Findings for the SAT-10 (Spring 2018 norms) are reported for 76 students in grades 2 through 8 and 10. The majority of students were Black (58%) and 40% of the students were first-time scholarship recipients (Table 2). Nearly all were eligible for free/reduced lunch subsidies (96%). No grade level had a sufficient number of students for reporting ($n < 25$), so scores were aggregated across all grade levels (Table 5).

Except for White students' Reading scores, the mean percentile scores for all subject areas and demographic groups were significantly below the 50th percentile. Comparisons between Black and White students' mean scores indicated that White students performed significantly higher in Reading. There were no gender differences in the mean percentile scores.

The SAT-10 Spring 2018 norms identifies a percentile score of 24 or greater as performing in an "Average" or "Above Average" Performance Cluster. The percentage of students meeting the minimum standard for "Average" is indicated in Table 4. Seventy-two percent (72%) performed in the Average range or higher for Reading and the percentage for Language was 65%. Math performance was somewhat lower, with 57% of students overall reaching the Average performance level. If students as a group were performing at the level of most students in the U.S., then it would be expected that 76% of students should be in the Average cluster or higher. In Reading, 76% of more females and White students meet this expectation and the other Reading scores are close to this mark. In Math, the percentage of students making the benchmark score is consistently lower than 76%. There were no significant race or gender differences in the percentage of students in the Average or Above Average performance clusters.

Table 5. Mean SAT-10 Percentile Scores and Performance Clusters for Grades 2-8, 10, and 11 (Spring 2018 Norms)

Grades	Group (N)	Reading		Language		Math	
		Mean Percentile	% Perf. Cluster	Mean Percentile	% Perf. Cluster	Mean Percentile	% Perf. Cluster
2-8, 10	All (73-75)	41*	72	37*	65	32*	57
	Black (42-43)	35*	68	36*	63	28*	48
	White (26-27)	50	78	39*	69	39*	70
	Female (45-46)	43*	70	37*	64	30*	51
	Male (29)	37*	76	37*	66	34*	66

% Perf. Cluster = Percentage of students that meet or exceed the Average Performance Cluster standard of a percentile score $\geq$ 24%.
* Mean score is significantly below the 50th percentile.

The Standardized Testing and Reporting Assessment (STAR)

STAR test results are reported for 31 students in grades 2 through 8 who took the test during the spring semester of the 2022-2023 school year. Female students comprised 71% of the sample. The racial background of students who took the STAR test differed from other tests (Table 2) in that White students were the largest group (42%). Forty-two percent were first-time scholarship recipients. All were free/reduced lunch eligible.

Table 6: Mean STAR Percentile Scores for Grades 2 – 8, 10, and 11

Grade	Group (N)	Reading	
		Mean Percentile	% Meeting Benchmarks
2-8	All (31)	42*	68

* Mean score is significantly below the 50th percentile.

Although STAR provides scores for Reading and Math, only Reading scores are presented in Table 6. Math scores are not included because only 24 test documents provided scores for Math. There were not enough students to disaggregate the scores by demographic groups. Mean scores for Reading were significantly below the 50th percentile. The benchmark for proficiency using the STAR Test is the 40th percentile. Most students met the benchmark for Reading (68%). However, as shown in Table 6, when compared to students in the nation who took the test, the mean score for scholarship recipients was lower than the 50th percentile of the nation.

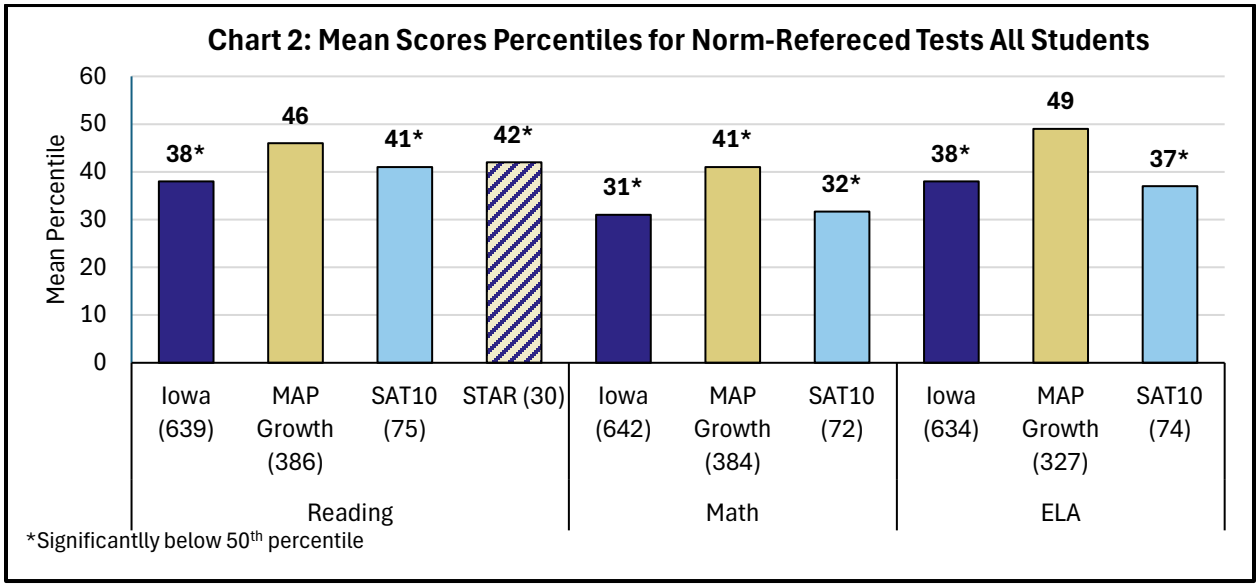
Summary of Norm-Referenced Test Results

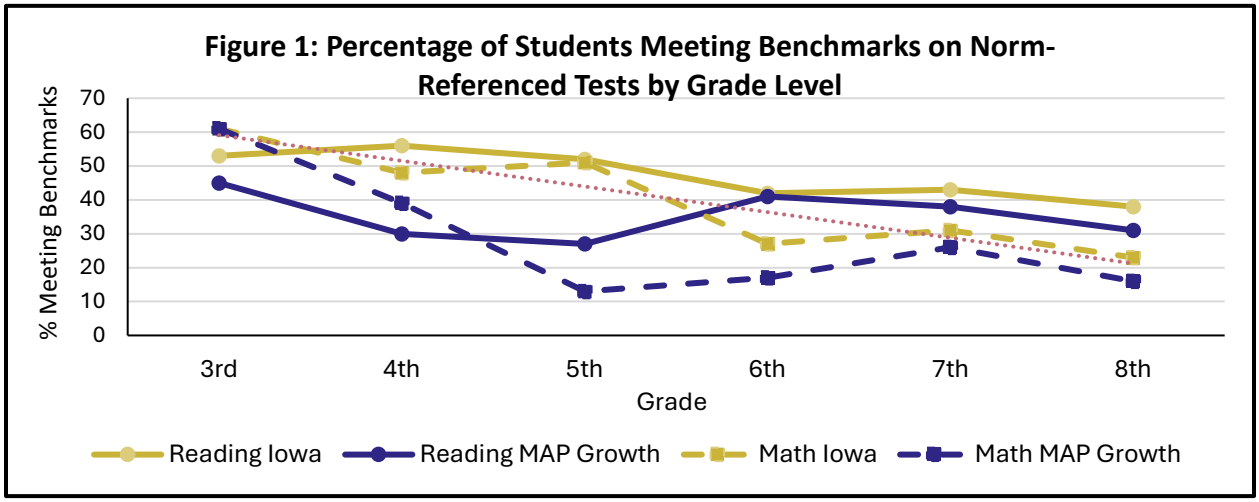
Although there was variability across the four norm-referenced tests, a pattern becomes evident when the results for each test are compared side by side. For norm-referenced tests, the 50th percentile score was used to indicate students' performance relative to national norms. As noted earlier, mean scores should be close to the 50th percentile if, as a group, scholarship recipients are achieving at levels similar to others in the U.S. Chart 2 graphically summarizes the mean percentile scores for students in all grades for each of the norm-referenced tests. The sample size associated with each test is in parentheses after the test name. Chart 2 shows that mean percentile scores were significantly below the 50th percentile, except for MAP Growth Language.

Academic achievement performance indicators on norm-referenced revealed that most students (> 50%) did not meet academic standards for their grade level on the two most commonly used tests, the Iowa Assessments and the MAP Growth (accounting for 1031 scholarship recipients), but on the two less commonly used tests SAT-10 (n = 76) and the STAR (n = 31) the majority of students met academic standards. The inconsistencies are likely due to differences in schools that utilize different tests or the

demographic make-up of the students who took these tests. Because the Iowa Assessments and MAP Growth results account for a larger portion of the scholarship recipients, these results are likely more representative of the academic achievement of the scholarship recipients. The analysis of the Iowa Assessments and MAP Growth results included grade-level achievement benchmark scores and the percentage of students meeting these benchmarks are plotted in Figure 1. In general, the percentage of students meeting achievement benchmarks appears to decrease from 3rd to 8th grade (Figure 1).

Together, the results generally indicate it is more typical for AAA scholars to perform below the median of students in the U.S., and the majority of students' scores fell below grade-level benchmarks set by testing agencies. Generally, the variability in findings across the tests suggests that there may be unmeasured factors associated with the schools using particular tests that could explain these results (e.g., school resources, class sizes, availability of help for struggling students).





2022-2023 Criterion-Referenced Test Results

Three tests are considered in this section: PSAT/NMSQT, PreACT, and ACT. In contrast to the norm-referenced tests, which focused on national percentile scores, performance on criterion-referenced tests focuses on proficiency benchmarks that indicate how well students perform compared to national grade

level standards. Table 7 provides the demographic information for students who took these three tests.

Table 7: Student Demographics for Criterion-Referenced Tests

Test	N	% Race				% Gender		% Poverty	% Years Receiving a Scholarship									
		Bk	Wh	His	Oth	F	M		1	2	3	4	5	6	7	8	9	10
PreACT	99	71	17	10	1	51	49	92	11	17	3	17	17	2	8	3	16	4
ACT	104	64	28	6	3	44	56	91	11	5	10	12	13	8	4	4	27	10
PSAT	88	40	25	31	3	57	43	93	10	11	0	11	9	3	6	12	28	9

Bk = Black; Wh = White; His = Hispanic; Oth = Race Other includes multiracial, unspecified, and other races and ethnicities (e.g., Asian, Middle Eastern)
% poverty is the percent eligible for free/reduced lunch

PSAT/NMSQT

Eighty-eight (88) scholarship recipients in the 8th, 10th, and 11th grades took the PSAT/NMSQT. The racial/ethnic make-up was comprised of Black, White, and Hispanic students. Most students (93%) were eligible for free/reduced lunch and 10% were first-time scholarship recipients. (See Table 7.) Results could only be presented for 10th and 11th grades due to the small number of 8th graders (n = 2). Scores could not be reported by racial/ethnic demographic groups because of the low number of students. There were enough female students to report their scores separately for 10th grade only.

The PSAT/NMSQT combines reading, writing, and language scores into an "evidenced-based reading-writing score" (EBRW). As a result, the combined scores are presented in Table 8. The EBRW and 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. The benchmark for EBRW corresponds to a scale score of 430 for 10th grade and 460 for 11th grade. For Math, the benchmark scores are 480 and 510 for 10th and 11th grades, respectively. Each of the mean scale scores presented in Table 8 is significantly higher than the corresponding benchmark score. Table 8 also presents the percentage of students meeting the PSAT/NMSQT benchmarks. For EBRW, the majority of students at each grade met the benchmark scores, and notably over 70% in 10th grade. For Math, the percentages were much lower, and the majority of students did not meet this standard. Together, the results provide a mixed picture for this group of students that varied with the subject area.

Table 8: Mean PSAT/NMSQT Scores and Percent Meeting Benchmarks for Grades 10 and 11

		Evidenced-Based Reading-Writing		Math	
Grade		Mean Scale Score	% Meets Benchmark	Mean Scale Score	% Meets Benchmark
10	All (46)	469**	72	443**	20
	Female (27)	491**	82	453**	22
11	All (40)	487**	58	437**	18

Reading-Writing benchmarks: 430 for 10th grade and 460 for 11th grade.
Math benchmarks: 480 for 10th grade and 510 for 11th grade.
** Mean score is significantly above the 50th percentile.

PreACT

PreACT test scores were included for 99 students in grades 7, 8, 10, and 11. Students who took the PreACT were predominantly Black (71%), 11% were first time scholarship recipients, and 92% were free/reduced lunch eligible. See Table 7 for detailed demographic information. There were only two students in each of

the 7th and 11th grades and only five students in the 8th grade who took the PreACT. Due to the small numbers in these grades, test scores are only reported for the 10th grade (N = 90).

For the PreACT, the critical scores are the scale scores (range 1-35) that correspond to the ACT college entrance exam scores, rather than percentile scores. Benchmark scores are provided to indicate college readiness. Specifically, according to the PreACT Technical Bulletin, these benchmarks indicate "the level of achievement required for students to have a 50% chance 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." Because the ACT is normally taken in the 11th grade, additional college readiness indicators are provided for 10th graders to account for the fact 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 for 10th grade are defined for each subject area: *In Need of Intervention, On the Cusp,* and *On Target.*

Table 9 presents the mean scale scores for 10th grade students and provides the corresponding college readiness indicator level. There was a sufficient number of students to report scores for Black students and for male and female students. The mean scale scores indicated that most students were "*On the Cusp*" of making the benchmark scores in Reading and in the "*Intervention*" group for Math. As a group, students were "*On Target*" for English, but Black and female students on average were "*On the Cusp*." There were no significant gender differences for any of the subject areas.

Table 9: Mean PreACT Scale Scores and Readiness Indicators for Grade 10

		Reading		English		Math	
Grade	Group (N)	Mean Scale Score	Readiness Indicator	Mean Scale Score	Readiness Indicator	Mean Scale Score	Readiness Indicator
10	All (90)	18	Cusp	15	On Target	16	Intervention
	Black (66)	17	Cusp	14	Cusp	15	Intervention
	Female (45)	18	Cusp	14	Cusp	16	Intervention
	Male (45)	18	Cusp	15	On Target	16	Intervention

Readiness indicators are for 10th grade students.

Table 10 present the percentage of 10th grade students who fell into each of the three readiness categories. For Reading and Math, the majority of students failed to meet benchmarks, but more than half of the students were "*On Target"* to meet the ACT readiness benchmarks for English across all demographic groups. The results for Math are noteworthy as 78% were "*In Need of Intervention*" for the group as a whole. For criterion-referenced tests, the goal is that 100% of students should meet benchmarks. The scores for the PreACT fall short of this ideal in all three subject areas.

Table 10: PreACT Percentage of Students in Grade 10 within Each Readiness Category

	Reading %			English %			Math %		
Group (N)	Inter-vention	On Cusp	On Target	Inter-vention	On Cusp	On Target	Inter-vention	On Cusp	On Target
All (90)	48	18	34	28	17	56	78	12	10
Black (66)	59	12	29	30	17	53	83	11	6
Female (45)	49	18	33	36	9	56	80	9	11
Male (45)	47	18	36	20	24	56	76	16	9

The ACT College Entrance Exam

ACT scores were reported for 104 students in 10th and 11th grade. The majority of this sample was Black (64%), and the free/reduced lunch rate was 91%. There were 11 (11%) first-year scholarship recipients. See Table 7 for demographic statistics. Only the 11th grade had a sufficient number of students to report scores, and there were enough students to break out scores for gender and Black students. See Table 11.

Table 11: Mean ACT Scores and Proficiency Rates for Grade 11

Grade	Group (N)	Reading		English		Math	
		Mean Scale Score	% Prof.	Mean Scale Score	% Prof	Mean Scale Score	% Prof
11	All (85)	18*	27	17	37	16*	9
	Black (57)	17*	19	16*	33	15*	7
	Females (39)	20*	36	17	41	16*	10
	Males (46)	17*	20	16*	33	16*	9

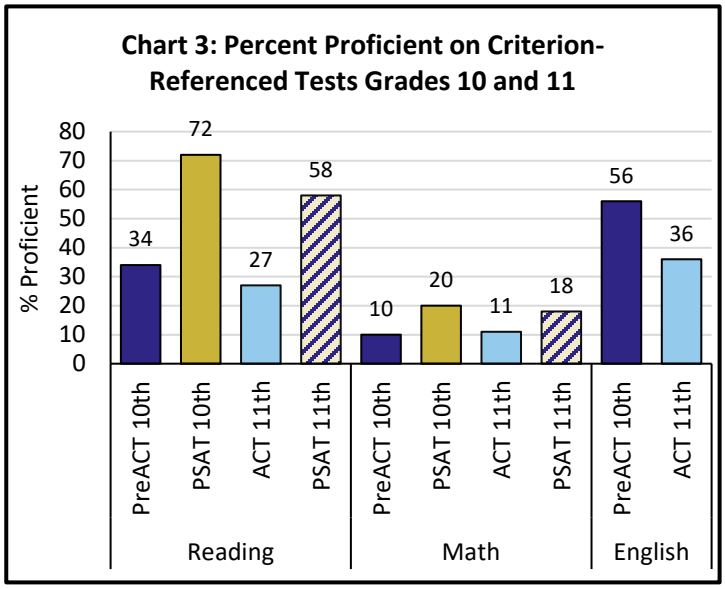
The benchmark scores for 11th grade ACT scores are 22 for Reading, 18 for English, and 22 for Math.
* Mean score is significantly below the benchmark

Similar to the PreACT, the relevant scores for the ACT are the scale scores (range from 1 to 36), which align with proficiency benchmarks for each grade level. The benchmark scores are similar to those for the PreACT and are interpreted the same way. The benchmark scores for 11th grade ACT scores are 22 for Reading, 18 for English, and 22 for Math. The average ACT scale scores were statistically significantly below benchmarks for college readiness for Reading and Math for all groups represented in Table 11. English scores for the group as a whole and females were not significantly different from the benchmark score of 18. Additionally, the percentage of students that reached proficiency was low for each subject, with the majority of students failing to reach proficiency, and these rates are especially low for Math.

Comparisons between genders revealed that the mean Reading score for 11th grade girls was higher than the mean score for boys. There were no other gender differences in mean scores for the other subjects, nor were there any differences in proficiency rates for any subject.

Summary of Criterion-Referenced Test Results

Chart 3 compares the proficiency rates for each of the criterion-referenced tests. A consistent finding across all tests and grade levels is that the majority of students failed to meet benchmarks in Math. Findings for English and Reading are mixed and depend on the test and grade level, but it seems that proficiency rates on the PSAT/NMSQT EBRW are much higher than those for Reading and English on the PreACT and the ACT. However, more reliable results are likely represented in the tests with larger sample sizes, specifically the PreACT grade 10 (n = 90) and ACT grade 11 (n = 85), compared to the PSAT/NMSQT, which had about half as many



students (*n's* = 46 and 40 for grades 10 and 11, respectively).

Objective 1 Conclusion

As a group, the results from the 2022-2023 school year are similar to those of the previous reports. Collectively, across the different tests and grades, the scholarship recipients generally fell below national norms in all subject areas. There is variability in student performance across tests and grade levels that show students performing closer to national norms or exceeding benchmarks. Generally, the variability in findings across the tests suggests that there may be unmeasured factors associated with the schools using particular tests that could explain these results (e.g., school resources, class sizes, availability of help for struggling students).

Summary for Objective 1: The Academic Achievement of Students in the Scholarship Program
⊙ Across all tests and grades, scholarship recipients generally scored below national norms in all subject areas. ⊙ Scholarship recipients tended to perform better in Reading and Language Arts/English compared to Math.
Findings for Norm-Referenced Tests
⊙ Scholarship recipients typically scored near or below the 50th percentile, with variations depending on the specific standardized test, grade level, and race. ⊙ Generally, less than half of the students met proficiency benchmarks on the Iowa, MAP Growth, and SAT-10, but on the STAR more than half met benchmarks for Reading. ⊙ In grades 3 through 8, the percentage of students meeting benchmarks in Reading and Math tended to decline as grade level increased. ⊙ White students generally performed better than Black students on the SAT-10.
Findings for Criterion-Referenced Tests
⊙ Math proficiency rates were lower than those for Reading and English across all tests, with 80% to 90% of scholarship recipients not meeting Math proficiency benchmarks. ⊙ On the PSAT/NMSQT and the PreACT, the majority of students met benchmarks for EBRW and English, respectively. ⊙ The majority of students failed to make benchmark scores for Reading on the ACT and PreACT and for English on the ACT. ⊙ Tests with larger sample sizes, such as the PreACT (10th grade) and ACT (11th grade) are likely to provide more reliable results.

Objective 2: Compare Scholarship Recipients to Alabama Public School Students

Students attending Alabama public schools in grades 2 through 8 took tests from the Alabama Comprehensive Assessment Program (ACAP), those in grade 10 took the PreACT, and those in grade 11 took the ACT college entrance exam. Although the ACAP has been in place for public school children since 2021, no direct comparisons can be made to the scholarship recipients because only one school with three students gave the ACAP test, which is not a sufficient number of students for a meaningful comparison. Additionally, Reading results were only available for 2nd and 3rd graders from ALSDE. Nevertheless, the ACAP results from ALSDE are reported and observations are made about how these results compare to the scholarship students. With respect to 10th graders, no results were publicly available for the PreACT for

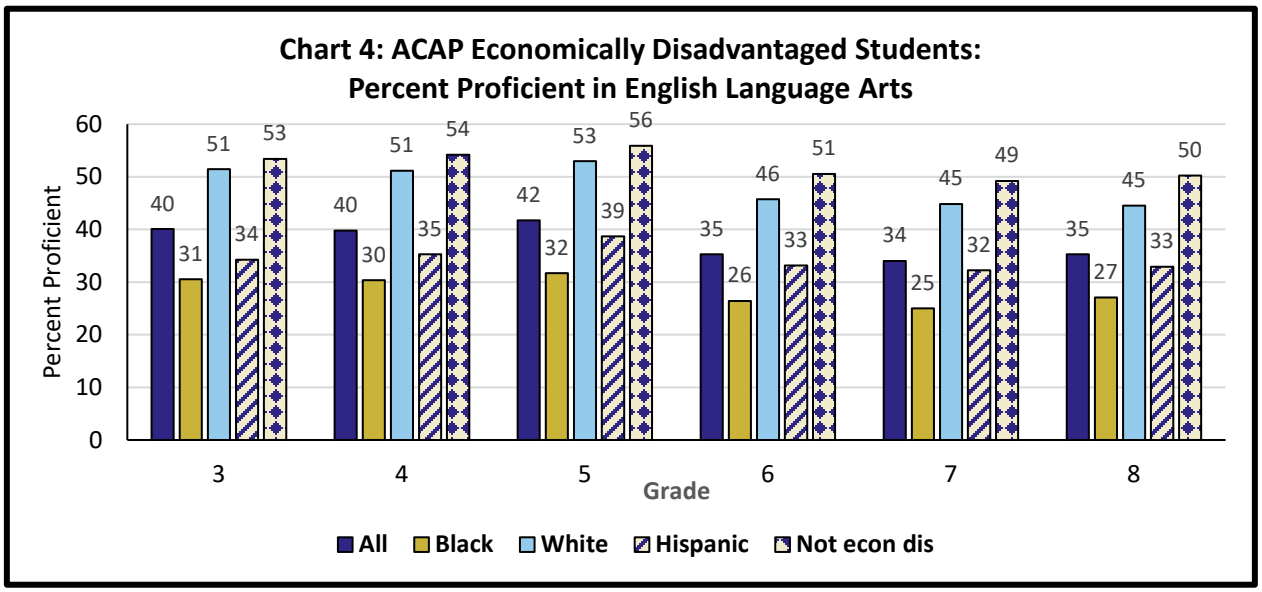
Alabama public school children. ACT scores were available for Alabama public school children in 11th grade, so comparisons can be made to the scholarship students.

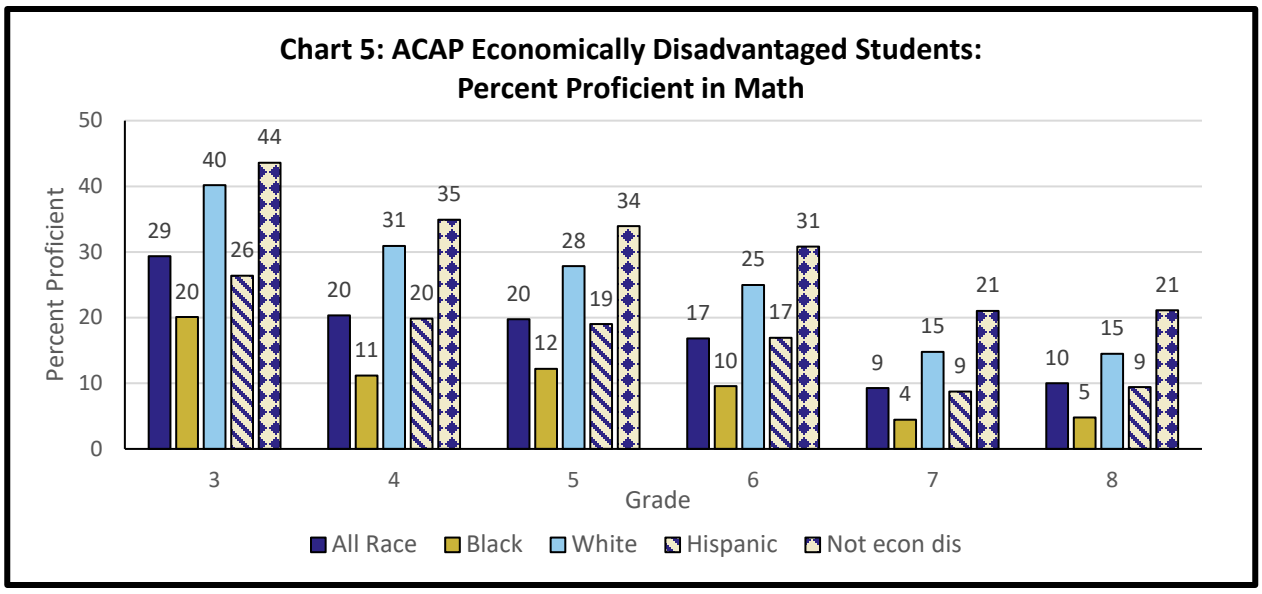
For Objective 2, economically disadvantaged public school students are the appropriate comparison group for scholarship students because 94% of the AAA scholarship students were eligible for free/reduced lunch. However, that demographic group was not available for all tests. Scores for Black, White, and Hispanic students are also reported when these groups were included in grade-level reporting for Objective 1.

The Alabama Comprehensive Assessment Program (ACAP)

Proficiency rates for English Language Arts (ELA) and Math were obtained from the ALSDE website, with scores aligned to the expected performance standards for each grade level. Charts 4 and 5 display the percentage of students in grades 3 through 8 who met the benchmarks for ELA and Math, respectively. Data are presented for all students (blue diamond bars), economically disadvantaged students (dark blue bars) and is further broken down by race. The results indicate that the majority of economically disadvantaged public school students did not meet proficiency benchmarks, with 58% to 66% failing to meet benchmarks in ELA and 71% to 91% not meeting the benchmarks in Math. In contrast, the blue-diamond bars show that when considering all public school students, approximately half met the ELA benchmarks, and 21% to 44% met the Math benchmarks. With respect to race, White economically disadvantaged students performed significantly better than their Black and Hispanic peers at every grade level. For Reading, test results for 2nd and 3rd graders were only available for all students combined, with no information for race or economic status. Based on data from ALSDE, 75% of second graders and 83% of third graders were reading at grade level.

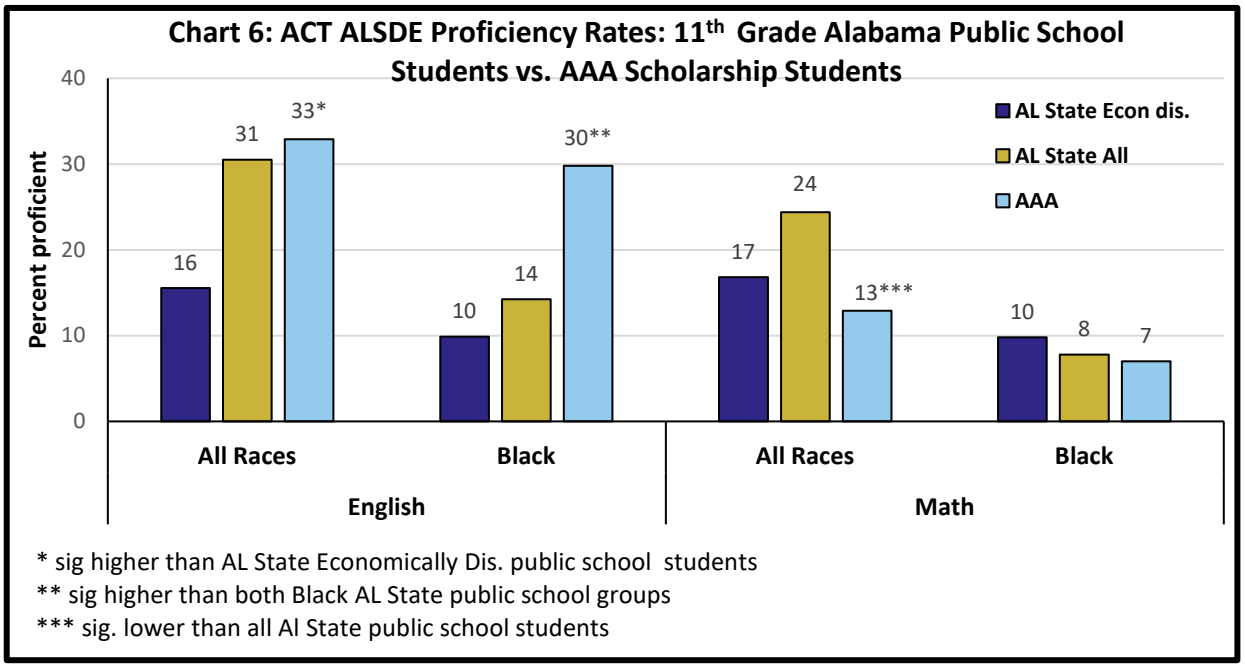
The performance of economically disadvantaged public school students is the most appropriate comparison group for scholarship recipients, but without a common test an accurate comparison cannot be made. Although proficiency rates for some subjects in specific grade levels on some tests may seem higher for one group or another, these rates cannot be compared because of fundamental differences among the tests and how they determine proficiency rates. Regardless of the test used to assess achievement, the majority of scholarship recipients and public school students from economically disadvantaged background fail to meet grade level performance standards.





ACT Scores Scholarship Recipients vs. Alabama Public School Students

The Alabama State Department of Education has established its own ACT proficiency levels for English and Math. The ALSDE does not report ACT Reading scores. For English the standard is a scale score of 19 or higher (the ACT uses 18), and for Math, the benchmark score is a scale score of 20 or higher (ACT uses 22). Chart 6 compares the proficiency rates of three groups: AAA scholarship recipients (light blue), Alabama public school students from economically disadvantaged backgrounds (dark blue), and all public school students, regardless of economic status (gold). A higher percentage of scholarship students were proficient in English compared to economically disadvantaged public school students, but not compared to all public school students. When examining Black students' proficiency rates for English, scholarship recipients scored significantly higher than Black students in both public school groups. Scholarship recipients had lower proficiency rates in Math than all public school students, but their scores were similar to those of economically disadvantaged public school students. Among Black students, there were no



significant differences in Math proficiency between the three groups. To summarize, contrasting results were found for the two subject areas. In English, the scholarship students performed similar to or better than public school students, depending on the public school comparison group. However, for Math, the scholarship recipients perform similar to or worse than the public school comparison groups, with no differences for Black students.

Objective 2 Conclusion

The goal of this objective is to compare the performance of scholarship recipients to that of public school children in Alabama. Unfortunately, until scholarship recipients are required to take the same tests as public school children, it will be impossible to draw robust conclusions for this objective. With respect to students in grades 3 through 8, it is clear the economically disadvantaged public school children failed to make grade level proficiency standards in English Language Arts and Math. By way of comparison, it appears that a similar statement holds true for scholarship recipients based on the metrics with which they were assessed. The limited data available for Reading indicates that 75% to 83% of public school students are meeting state standards in 2nd and 3rd grade, whereas the Reading proficiency rates for the scholarship students in these grade levels are much lower.

Stronger conclusions can be made for 11th grade students compared to the other grades required to take standardized tests because both groups of students took the ACT. Findings for ACT English scores indicated that scholarship recipients exceeded their economically disadvantaged public school counterparts, and their performance was comparable to all students in the state (regardless of economic status and race). Black scholarship recipients performed much better than Black students in the public schools for English. For Math, performance of scholarship students and economically disadvantaged public school children are comparable, but public school children as a whole had higher proficiency rates. As noted in our previous reports, because 11th grade students are near the end of their state mandated education and the majority of scholarship students who took the ACT had received a scholarship for five years or more, these results might represent the cumulative effects of receiving a scholarship. The findings for this report suggest that economically disadvantaged 11th graders performed better on ACT English test if they had received a scholarship, but that any advantage was lost when Math scores were considered. Across both subject areas and all comparison groups, the majority of all students who took the ACT failed to meet benchmark standards. Overall, similar to past reports, because these comparisons include just a small percentage of the scholarship students, some caution must be taken in generalizing them to the larger group of scholarship students. Only 42% of the 11th graders who received scholarships took the ACT. An important consideration for the ACT results is that the schools that administer the ACT may differ from other scholarship recipient schools in their curriculum being geared toward college preparation.

Summary for Objective 2: Scholarship Recipients vs. Alabama Public School Students
⊙ **Due to the lack of appropriate comparative data, strong conclusions cannot be made** for the relative performance of the scholarship recipients and the scholarship recipients.
ACAP findings for 3rd through 8th grade: ⊙ The majority **of economically disadvantaged public school children did not meet proficiency standards for Language Arts and Math.** • **White** economically disadvantaged students **performed relatively better than other racial groups.** ⊙ Performance **in Language Arts was better than performance in Math**. ⊙ **Most 2nd and 3rd graders met Reading proficiency standards.** ⊙ **No direct comparisons could be made between Alabama public school students and the scholarship recipients** because only a few scholarship students took the ACAP.
ACT findings for 11th graders: ⊙ The majority of both scholarship recipients and public school **children failed to meet benchmarks scores for both English and Math.** ⊙ The **proficiency rates were higher for scholarship recipients in English compared to** economically disadvantaged public school students. ⊙ There were **no differences between scholarship recipients and economically disadvantaged public school students in proficiency rates for Math.**

Objective 3: Changes in Achievement across Time

Changes in scholarship students' performance over time are examined in this section of the report. Specifically, data are analyzed to assess if longer participation in the AAA program is related to higher standardized achievement scores. Changes in scholarship students' scores are compared to changes in public school students' scores over the same time period to distinguish between changes that are the result of participating in the AAA program versus general societal changes that affect all students. As noted in previous reports, there are two challenges meeting this objective. First, ideally, such an analysis would calculate the average change in national percentile scores or proficiency groups over time for scholarship students and compare it to comparable changes for public school students. However, change in scholarship students' performance from one year to the next is difficult to assess for several reasons:

- **Inconsistent testing:** Many students do not take the same test each year due to schools changing tests, students changing schools (especially from 8th grade into high school), or no test data being available (because a student was not required to test due to his or her grade or the test report was not submitted). Thus, a large percentage of students would be excluded from this longitudinal analysis.
- **COVID-19 disruptions**: The pandemic further complicated this analysis by disrupting annual testing in 2020.
- **Test changes in public school testing:** ALSDE changed the required achievement test for grades 2 through 8 from the Scantron Performance Series to the ACAP in 2021, so there are only a few years of ACAP data for comparison.

Second, as noted throughout this report, without a common test across the two groups of students, definitive comparisons cannot be made. Combined with the inconsistent testing described above, the amount of data that can be included in analyses is limited. However, 11th graders in both public and scholarship receiving

schools took the ACT. Thus, a direct comparison in performance over time could be made for this test for a small number of students.

With these limitations in mind, we follow previous reports and take two approaches to examine change over time. The first examines the relationship between the number of years a student had received a scholarship, and their achievement test scores for the 2022-2023 academic year. Correlation analyses were conducted between test scores and years in the scholarship program using the test data included in Objective 1. These correlation analyses include the greatest number of scholarship students and test types. Second, the 2022-2023 cohort of AAA students had relatively more years attending a scholarship school other than their assigned public school compared to earlier time points. If the AAA program is having a positive impact on achievement, then it might be expected that high school scholarship recipients in more recent years should have higher scores compared to earlier cohorts. Because the ACT has been consistently administered to 11th graders over the years, performance was compared between scholarship and public school students over seven years. If participation in the AAA is the cause of improving scores, the same level of improvement should not be evident for the economically disadvantaged public school children.

Correlations between 2022-2023 Test Performance and Number of Years Receiving a Scholarship

To assess if there is a relationship between performance on the 2022-2023 achievement tests and the number of years a student was in the scholarship program, a series of correlation analyses were conducted. Correlations can be positive, negative, or not significant, and they can range from -1 to +1. They reveal the direction of change over time (i.e., increasing or decreasing), but do not reveal the amount of change over time. A significant positive correlation would indicate that the longer a student was in the scholarship program, the better they performed on the achievement tests. A significant negative correlation indicates the opposite. Significant correlations, however, cannot be interpreted as participation *causing* scores to change; rather they can only suggest that the two are related. Non-significant correlations suggest that there is no relationship between achievement test scores and the number of years a student had received a scholarship.

Similar to making comparisons based on mean scores or proficiency groups, a minimum sample size is necessary to detect a reliable correlation. A minimum sample size of 60 is necessary to detect a moderate relationship between test performance and the number of years receiving a scholarship. Additionally, only students in grades 6 or higher were included because grades lower than that had a more restricted range for the number of years a student could have received a scholarship. For example, a student in 2nd grade could have at most three years of participation in the AAA program (kindergarten, first, and second grades); whereas 11th graders might have up to 10 years of participation. A restricted range can cause correlations to be attenuated and thus not provide an accurate picture of the relationship between variables. As a result, correlation analyses were only conducted for four tests that had 60 or more students in grade 6 or higher: Iowa Assessment, MAP Growth**,** PreACT, and the ACT tests. These analyses represent 625 of the students with available test data. In these analyses, national percentile scores were used to measure test performance for the Iowa Assessment and MAP Growth, and scale scores were used for the PreACT (10th grade) and ACT (11th grade).

Correlations calculated between the number of years a student had received a scholarship (one to ten years) and their percentile scores in Reading, English/Language, and Math for the Iowa Assessment and MAP Growth yielded only two significant positive correlations, both for the Iowa Assessment:

- Iowa Assessment Language: $n = 376$, $r = .148$, $p = .004$
- Iowa Assessment Math: $n = 384$, $r = .112$, $p = .028$

Next correlations were calculated between the number of years a student had received a scholarship and the scale scores for 10th graders on the PreACT and 11th graders on the ACT in the three subject areas. None of these correlations were significant.

To summarize, only two out of the 12 correlations calculated were significant. These positive correlations suggest that students who took the Iowa Assessment may improve in some subject areas the longer they participate in the program, but for all other tests, there was no relationship between years of participation and academic achievement. The significant correlations are relatively small (possible range -1 to +1), which indicates that performance on the Iowa Assessment is only weakly related to the number of years students had participated in the scholarship program.

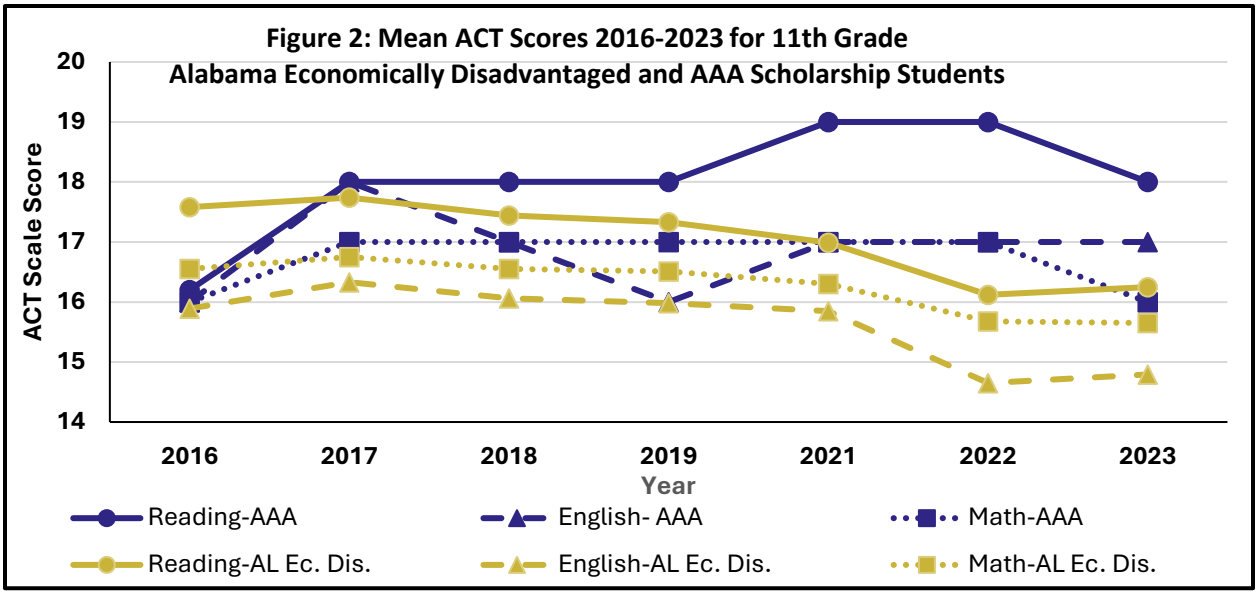
Comparison of Scholarship Recipients to Alabama Public School Children in Grade 11 on the ACT over Time

The 2022-2023 cohort of AAA students had relatively more years attending an alternative school rather than their assigned public school compared to earlier time points. If the AAA program is having a positive impact on achievement, then it might be expected that scholarship recipients in more recent years should have higher scores and rates of proficiency compared to earlier cohorts. If participation in the AAA is the cause of improving scores, the same level of improvement should not be evident for the economically disadvantaged public school children.

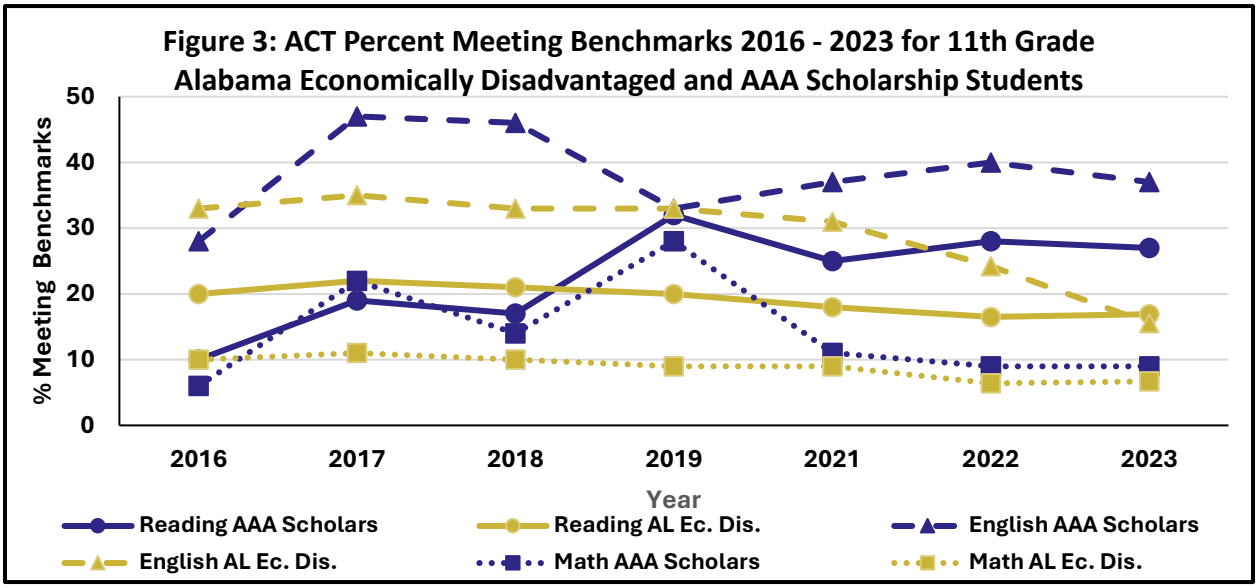
Mean ACT scores for 11th grade were gathered for the scholarship students from previous reports starting in the 2015-2016 academic year through 2022-2023. Scores are not included for 2020 because many students were not tested due to the COVID-19 pandemic. Comparable data were available from PARCA for Alabama economically disadvantaged public school children.

Figure 2 plots the mean ACT scale score for Reading, English, and Math for AAA Scholarship students (in blue) and public school students (in gold) and Figure 3 plots the corresponding proficiency rates. It should be noted that often seemingly large changes in scores in the Figures may not be statistically significant. The non-significant statistical tests tell us that, despite their size, the differences between the groups are not reliable. When this happens, it is usually related to the relatively small sample sizes being compared for some of the years.

Statistical analyses for each subject area examined whether the mean scores for AAA Scholarship students are improving over time (Figure 2). In the 2022 report, we reported that that ACT scores had not significantly improved since 2016 with the exception of Reading, and the same is true in the current report. Follow-up analyses indicated that the mean Reading scores for 2019 and 2021 through 2023 were significantly higher than those for 2016, but not higher than those in the intervening years (2017 and 2018) nor were there differences among the 2019 through 2023 scores in Figure 2. By comparison, the mean scores for the disadvantaged public school children varied only slightly over time, rarely changing more than one scale score point from year-to-year over the seven years represented. Post-COVID-19 scores for 2021 and 2022 declined in all subject areas but leveled off in 2023. Additional comparisons revealed that starting in 2021, the mean ACT Reading scores for the scholarship recipients were significantly higher than the economically disadvantaged public school students. In 2022 and 2023, the English and Math scores were also significantly higher for the scholarship recipients.



The proficiency rates plotted in Figure 3 provide further insight in to changes in scholarship recipients scores over time. It should first be noted that in any given year, less than 50% of the students met the benchmarks for Reading, English, or Math for both groups of students. In the 2022 report, which included data through 2021, it was noted that even when there was a statistically significant improvement between earlier and subsequent years, the improvements for scholarship students were not maintained. From 2021 to 2023 scores have leveled off, showing no significant changes among these years. Similarly, scores for public school children also changed very little from 2021 through 2023, except for English scores, which are declining. Scholarship recipients had significantly higher proficiency rates than economically disadvantaged public school students for Reading and English in 2022 and 2023.



Objective 3 Conclusion

Together, the data on scholarship students does not reveal a consistent pattern of improvement or decline over time. The correlation analyses indicated that the number of years that a student had been in the AAA program was not strongly related to achievement test scores, supporting conclusions from previous reports. Analysis of ACT scores also does not indicate positive increases in scores over the years. ACT data from 2022 and 2023 indicate that in Reading and English scholarship students might be performing better than their economically disadvantaged public school counterparts. However as noted in Objective 2, a strong statement about the performance of two groups of 11th graders cannot be made because less than half of the 11th grade scholarship recipients (42%) took the ACT, in contrast to all of the 11th graders in public schools. As noted throughout this report, there are unmeasured factors that influence which tests schools administer to their students and these can impact test outcomes. Because the ACT is a college entrance exam, it is possible that the scholarship recipients who took this exam had expectations to earn a higher degree than other scholarship recipients or attended schools that provided more preparation for college.

Summary for Objective 3: Changes in Achievement across Time
⊙ **The number of years of participation in the scholarship program was not strongly associated with significant improvement** on standardized tests scores. ⊙ Over the same time period, **economically disadvantaged public school students** in Alabama also did not show consistent improvement on the ACAP or the ACT.
⊙ The **number of years** that a student participated in the scholarship program was generally **not strongly correlated with higher** achievement test scores. ⊙ **With the exception of Reading,** ACT scores have not significantly improved since 2016. ○ Reading scores in 2019 and 2021 through 2023 were significantly higher than those for 2016, but scores from 2018 on did not differ from each other, suggesting that students on average are not gaining in reading achievement. ○ A strong statement about improvements on the ACT over time cannot be made because only 42% of the 11th grade scholarship students took the ACT.

General Conclusion

The current report is the fifth report since the inauguration of the AAA program in the 2013-2014 academic year and provides the most recent assessment of how the scholarship program enacted through the AAA affects the academic achievement of scholarship recipients. The academic performance of scholarship recipients was analyzed by utilizing the demographic and test score data provided annually to the SGOs by the schools that enroll students with scholarships. Many factors that impact the reliability and validity of the findings were noted throughout the report, and these are nearly all linked to the lack of a common test among schools. Within these limitations, the evaluation addressed three objectives.

- Objective 1 reported on the academic performance of the AAA scholarship recipients for the 2022-2023 school year. Scholarship recipients generally scored below national norms, performing better in Reading and Language Arts/English than Math. Most AAA scholarship recipients scored at or below the 50th percentile on norm-referenced tests, and fewer than half met proficiency benchmarks. Performance in Reading and Math decreased as students advanced from grade 3 to 8. White students typically outperformed Black and Hispanic students on the SAT-10. Criterion-referenced tests revealed that 80% to 90% of students did not meet Math proficiency standards. Findings for English and Reading are mixed and depended on the test and grade level. Although proficiency rates on the PSAT/NMSQT EBRW indicated that the majority of students were proficient, more reliable results are likely represented on the PreACT and ACT, which had larger

sample sizes. On the PreACT and ACT, AAA scholarship students showed low proficiency in Reading. Performance in English was better on the PreACT, as nearly half of scholarship recipients were proficient.

- Objective 2 compared the scholarship students to economically disadvantaged Alabama public school students. No direct comparisons could be made for elementary and middle school students due to lack of a common test. On the ACAP, the majority of economically disadvantaged public school children failed to meet benchmark scores. By way of comparison, it appears that a similar statement holds true for scholarship recipients based on the metrics with which they were assessed. A better comparison was made for 11th graders on the ACT since both groups of students took this test. The ACT proficiency rates were higher for scholarship recipients in English compared to economically disadvantaged public school students, but there were no differences in proficiency rates for these groups in Math. While the ACT results show some advantages to scholarship recipients, they must be viewed with caution because less than half of the 11th grade scholarship recipients took the ACT, and there may be differences in the high school curriculum of schools that administer the ACT compared to high schools that do not.
- The third objective assessed if scholarship recipients' achievement scores improved over time. Based on correlation analyses, which included the greatest number of scholarship recipients, the number of years that a student had been in the AAA program was not strongly related to achievement test scores. This finding supports conclusions from comparable analyses in the 2022 report. Similarly, analysis of ACT mean scores and proficiency rates does not indicate positive increases in scores over the years. However, the *mean* ACT Reading scores for the scholarship recipients were significantly higher than the economically disadvantaged public school students starting in 2019. In 2022 and 2023, the English and Math scores were also significantly higher for the scholarship recipients. *Proficiency rates* in Reading and English were higher for scholarship students in 2022 and 2023 compared to economically disadvantaged public school children. These trends will need further study in future reports.

A consistent conclusion in the evaluation of the AAA since 2016 is that the majority of AAA scholarship students performed similarly to their peers in public schools in that they often fell below national expectations for their grades. Students in both groups tended to perform better in Reading and English/language arts than Math. It is important to note that a nationally representative longitudinal study of academic performance conducted by the NAEP revealed that achievement tests scores in Reading and Math have been generally stagnant or slightly declining since 2012 (Walton, 2023). Performance on the ACT was a bright spot for the AAA program, but interpretation of this as indicating that the scholarship recipients are more successful than public school children is problematic because the scholarship students who took the ACT and the schools that administered it might not be representative of all scholarship recipients. If the AAA scholarship program was having a strong impact on students' academic achievement, then we would expect higher proficiency rates and mean percentile scores that compare more favorably to national norms across all grade levels and tests, as well as a strong correlation between test scores and the number of years in the program. However, our analysis found that only two out of 12 correlations were significant, and even those were weak ($r \leq .15$). This suggests that factors other than the number of years in the scholarship program contribute to improved proficiency in Language and Math on the Iowa Assessment.

Limitations

Since the initial report in 2016, the lack of a common assessment has been noted as a critical limiting factor in drawing strong conclusions regarding the academic achievement of scholarship recipients relative to students attending public schools. Only a few scholarship students took the ACAP, and although more scholarship students could be directly compared to public school children on the ACT (N = 85 11th graders),

they only represented 4% of the AAA students who were required to test. An accurate model of the effects of the scholarship program would require statewide student-level assessments that use the same standardized test and link test scores to student demographic information.

This report made the best use of the data available, but there are challenges inherent in working with the data from multiple standardized tests that make it difficult to draw conclusions. For example, students' performance sometimes appears better on one test compared to another. However, as noted in the previous reports, the results for a particular test are confounded by idiosyncratic characteristics of the schools that use that test, such as a particular pedagogical model, accreditation agency, or demographic characteristics such as the composition of race, household income, or number of years receiving a scholarship. These confounding factors cannot be readily accounted for in the evaluation.

Small sample sizes for some tests and missing data (see Flow Chart p. 5) also impact the statistical reliability and validity of the report. Some schools opted to evaluate student performance using tests with outdated national norms, which may save money for a school, but the value of this practice for evaluating student learning is questionable. Several schools provided test scores for assessments taken in the fall, which as noted in Objective 1, do not represent the learning a student has achieved during the year.

Finally, it is important to recall that the AAA scholarship program targets low-income students and has been utilized by families belonging to demographic groups (e.g., racial minorities) that have historically lagged behind others in the state and the U.S. in academic achievement. The results presented here are aggregated across all such scholarship students. Although the findings of this report do not suggest large advantages to scholarship students, some students might be performing very well compared to their public school counterparts; however, these students' performance is offset by others who are not performing as well. This report, along with state and national data, make it evident that sustained and lasting improvement for low income students is difficult to achieve.

Sources

Alabama Department of Education (Report Card) Accountability Indicators. https://reportcard.alsde.edu/SupportingData_Accountability.aspx

Alabama Achieves. https://www.alabamaachieves.org/

College Board/National Merit Scholarship Corporation. PSAT/NMSQT: Understanding Your Scores 2023-24. https://satsuite.collegeboard.org/media/pdf/psat-nmsqt-understanding-scores.pdf

National Norms for PreACT Reported Scores During 2022-2023 Reporting Year. ACT success. https://success.act.org/s/article/PreACT-and-PreACT-89-US-Ranks

Public Affairs Research Council of Alabama. ACT 2015-2022. https://public.tableau.com/app/profile/parca/viz/ACT2015-2022/AlabamaACTResults

PreACT Interpretive Guide for Student and Aggregate Reports. ACT success. https://success-stress.act.org/s/article/Interpretive-Guide-for-Student-and-Aggregate-Reports

Sawyer, R. (1987). Determining minimum sample sizes for estimating prediction equations for college freshman grade average. American College Testing Program.

Tran, S. T., Majumdar, R., Hu, A., & Meyer, J. P. (2022). Default Cut Scores for MAP Growth Reading and Mathematics Assessments. https://www.nwea.org/uploads/2022/07/MAP-Growth-Default-Cut-Scores_NWEA_linking-study.pdf

Walton, E. (2023). Performance declines in basic mathematics and reading skills since Covid-19 pandemic are evident across many racial/ethnic groups. *NAEPPLUS+: The official blog of the nation's report card*. https://nces.ed.gov/nationsreportcard/blog/pandemic_performance_declines_across_racial_and_ethnic_groups.aspx