



PROGRAM EVALUATION REQUIREMENTS FOR OPPORTUNITY SCHOLARSHIPS

Report to
the Joint Legislative Education Oversight Committee

Submitted by
The North Carolina State Education Assistance Authority

March 1, 2023

As Required by
Section 8A.8(a) of S.L. 2022-74

Background

The North Carolina State Education Assistance Authority (SEAA) is a political subdivision of the State created under Article 23 of Chapter 116 of the NC General Statutes to help North Carolinians pay for education. The work of SEAA includes administering various student financial aid programs. Among these programs are two State-funded scholarship programs for eligible K12 students¹ to pay tuition and fees to attend nonpublic schools in North Carolina - the Opportunity Scholarship Program (“OSP”) and the Education Student Accounts for Children with Disabilities Program (“ESA+ Program”) (collectively the “K12 Programs”).

In 2022, the North Carolina General Assembly directed SEAA, in collaboration with the Department of Administration, Division of Nonpublic Education, and the Department of Public Instruction, to report to the Joint Legislative Education Oversight Committee on the following information:

- (1) Options and a timeline to implement the recommendations of the March 1, 2018, report of the task force established pursuant to Section 10A.6 of S.L. 2017-57.
- (2) The estimated cost of each option provided pursuant to subdivision (1) of this subsection.
- (3) Any legislative recommendations on improving the evaluation of students receiving scholarship grants pursuant to Part 2A of Article 39 of Chapter 115C of the General Statutes.

The March 1, 2018, Task Force Report, entitled “Study of Opportunity Scholarship Student Evaluations”, recommended the use of a common standardized test administered to certain Opportunity Scholarship recipients attending nonpublic schools and to certain public school students to be the most effective, valid, and reliable evaluation option in order to study the learning gains and losses of students in the Opportunity Scholarship Program. As required by Section 8A.8(a) of Session Law 2022-74, this report sets forth the background, options and recommended timing of the implementation of the March 1, 2018, task force report, as well as SEAA’s legislative recommendations on improving the evaluation of students receiving the Opportunity Scholarships.

Statutory Requirements

K12 Program Annual Testing Requirements

Nonpublic schools with K12 scholarship students² are required to administer “a nationally standardized test or other nationally standardized equivalent measurement selected by the chief administrative officer of the nonpublic school to all eligible students whose tuition and fees are paid in whole or in part with a scholarship grant enrolled in grades three and higher.”³ The only requirements regarding the type of test selected is that (1) for grades three through eight, the nationally standardized test (or other equivalent measurement selected) must measure achievement in the areas of English grammar, reading, spelling, and mathematics and (2) for grades nine through 12, the nationally standardized test

¹G.S.115C-562.1(3).

² If a nonpublic school is a G.S. 115C-562.5 compliant school and receives funds through direct disbursement. See G.S. 115C-595(a1).

³ G.S. 115C-562.5(a)(4).

(or other equivalent measurement selected) must measure either (i) achievement in the areas of English grammar, reading, spelling, and mathematics or (ii) competencies in the verbal and quantitative areas. There is no mandate that a certain type of standardized test be selected by the nonpublic schools.

Nonpublic schools are required to submit individual test scores to the Authority on an annual basis no later than July 15.⁴ In addition, schools with more than 25 K12 Program scholarship students are required to “report to the Authority on the aggregate standardized test performance of eligible students.”⁵

The K12 Programs’ Learning Gains and Losses Report

SEAA is to select an independent research organization to study the learning gains or losses of students receiving the K12 scholarship grants.⁶ The learning gains and losses report is to be submitted to the Joint Legislative Education Oversight Committee by the SEAA on an annual basis.⁷

The selected research organization is charged under the Program statute with preparing a report for SEAA on the following:

- (1) Learning gains or losses of students receiving scholarship grants. The report shall include learning gains of participating students on a statewide basis and shall compare, to the extent possible, the learning gains or losses of eligible students by nonpublic school to the statewide learning gains or losses of public school students with similar socioeconomic backgrounds, using aggregate standardized test performance data provided to the Authority by nonpublic schools and by the Department of Public Instruction.
- (2) Competitive effects on public school performance on standardized tests as a result of the scholarship grant program. The report shall analyze the impact of the availability of scholarship grants on public school performance on standardized tests by local school administrative units to the extent possible, and shall provide comparisons of the impact by geographic region and between rural and urban local school administrative units.

Task Force Report

In 2017, at the direction of the General Assembly, SEAA established a task force (the “Task Force”) to study the evaluation of students receiving Opportunity Scholarships. The Authority collaborated with the Department of Administration, Division of Nonpublic Education, and the Department of Public Instruction. In addition to representatives from the three collaborating organizations, Task Force members included representatives from nonpublic schools of varying sizes, organizations representing parental school choice (specifically Parents for Educational Freedom in North Carolina), organizations

⁴ G.S. 115C-562.5(a)(4).

⁵ G.S. 115C-562.5(c). However, 115C-562.5(c) does not apply for the 2022-23 school year and schools will not be deemed ineligible for scholarship funds due to its inability to provide the aggregate standardized test performance of its scholarship students for the 2022-23 school year. See S.L. 2022-74, Section 8A.8(b) and(c),

⁶ G.S. 115C-562.7(c).

⁷ For the 2022 and 2023 calendar years, no learning gains and losses report was required to be submitted to the legislature and the requirements of G.S. 115C-562.7(c) do not apply to those years. See S.L. 2022-74, Section 8A.8(d).

representing nonpublic schools, two independent research organizations, and a representative from a local public school system, as required by the General Assembly.

Under Session Law 2017-57, Section 10A.6.(b), the Task Force was charged with:

studying the most effective, valid, and reliable method of evaluating learning gains or losses of students receiving scholarship grants and comparing the learning gains or losses of those students to public school students with similar socioeconomic backgrounds, including the potential for adoption of a nationally normed common test for students participating in the evaluation. In doing so, the task force shall also consider the most reliable manner of establishing causal relationships to student performance outcomes while achieving minimal interference with the operation of the participating nonpublic and public schools, including limited sampling and other suitable research design methods.

The Task Force met to consider the challenges to and possible methodologies for comparing learning gains or losses between participants in the OSP and selected North Carolina public school students. On March 1, 2018, the Task Force issued its report, a copy of which is attached hereto as Attachment 1. The Task Force determined that meeting the statutory charge to evaluate student outcomes via “the most reliable manner of establishing causal relationships” is not possible with standardized testing data as currently required in the Opportunity Scholarship Program.

Most laypeople would expect one standardized test to equal another. However, as part of its work, the Task Force clarified that reliable test concordance (the degree of agreement between two tests) can be established only via a well-designed study. New versions of standardized tests require updated concordance studies. Therefore, comparing results among a variety of standardized tests does not produce reliable data, and establishing concordance among those tests in order to compare the results is a difficult and expensive undertaking.

The Task Force also addressed the challenges for researchers regarding establishing a causal relationship, which is a high bar for an evaluation. The recommendation in 2018 was for evaluators to sample students enrolled in a nonpublic school who receive the Opportunity Scholarship, and to also sample students who remain in public school. Both groups would take the same nationally normed test, which circumvents the challenge of establishing concordance. Students in the study sample would take the test at least twice: once during the Opportunity Scholarship application process and again after one year, to assess academic growth.

The Task Force noted that the cost of tests, third-party administration, and operationalizing a causal study in an education environment are significant. An appropriation included in potential legislation would prevent a financial burden on participating private and public schools.

Legislative Recommendations on Improving the Evaluation of Opportunity Scholarship Students

The evaluation design as recommended by the 2018 Task Force would be the most effective, valid, and reliable evaluation option, and would provide a comparison of learning gains or losses between participants in the Opportunity Scholarship Program and public school students.

In order to conduct the evaluation as recommended by the Task Force, the following legislative changes would be necessary:

- Add a requirement that students receiving an Opportunity Scholarship award may be selected to take a common test (“the Common Test”) through sampling for evaluation purposes and make this requirement a condition of receiving the Scholarship.
- Add a requirement that nonpublic schools administer or allow the administration of the Common Test to any scholarship students who are selected to take the Common Test in order for the school to maintain its eligibility to directly receive Opportunity Scholarship funds on behalf of students.
- Add a requirement that the State Board of Education direct the local school administrative units (LEAs) to administer the same Common Test to a certain percentage of students attending public schools.
- Direct the independent research organization conducting the evaluation to select the Common Test or, in the alternative, make a recommendation to the Authority as to which Common Test should be used, and use the smallest sampling size possible in the evaluation.
- Permit nonpublic schools and public schools to submit test results from any students who have already taken the Common Test at their respective schools during the relevant school year and deem those test results sufficient for those students.
- Require that the State Board of Education communicate directly with the independent research organization regarding the testing data, including establishing procedures for the transmission of the student data directly to the research organization.
- Require that the Authority communicate directly with the independent research organization regarding the testing data, including facilitating the transmission of the student data from the nonpublic schools directly to the research organization to ensure protection of personally identifiable student information.
- Allow at least a year for SEAA to select the research organization and structure the study. The testing could then begin, with the results not available until the first cohort of students were tested twice.
- The cost of the evaluation would depend on the frequency of reporting. Currently, the report is required on an annual basis. See “Initial Implementation Timeline and Costs” section for additional information.

Additional Recommendations Relating to the Collection and Submission of Opportunity Scholarship Student Test Scores

SEAA would also recommend legislative changes related to the submission and collection of test scores by the schools.

1. Modify Schools’ Requirement to Submit Individualized Test Scores to NCSEAA.

Currently, participating nonpublic schools are required to submit individualized test scores for each Opportunity Scholarship recipient in third grade or higher to the Authority. Due to the increasing volume of student test scores (approximately 16,000 Opportunity Scholarship students are required to be

tested for the 2022-23 school year)⁸, SEAA’s administrative role in the collection of the test scores has become increasingly burdensome. Instead, SEAA proposes that schools be required to retain test scores for each Program recipient in their records and on their premises. Participating schools would be required to certify compliance with the testing requirements on an annual basis and make documentation available to SEAA for production or inspection upon request. SEAA may request test scores for compliance reasons when deemed necessary, including for specific students selected for testing in conjunction with the learning gains and losses report and turn those over to the research organization.

In addition, this proposed change would result in uniform recordkeeping requirements for nonpublic school testing that are overseen by SEAA and DNPE. Nonpublic school law also requires that nonpublic schools administer a nationally standardized test or its equivalent to *all* students in grades three, six, nine, and eleven each year.⁹ State law requires that those test results be kept on file at the school’s office for one year after the testing date for annual inspection by DNPE. We would request that G.S. 562.5(a)(4) be amended to provide that the participating nonpublic schools should keep all individual test scores on file at their school and available for inspection or production, rather than submit the individual test results to the Authority.

2. Schools’ Requirement to Submit Aggregate Test Scores.

Nonpublic schools with more than 25 Opportunity Scholarship recipients are required to report to SEAA on the “aggregate standardized test performance of eligible students” in grades 3 and higher on an annual basis. However, experts in psychometrics interviewed for the Task Force study, as well as upon consultation with SEAA on aggregate standardized test performance data reporting, have indicated that aggregate scores of an isolated portion of students in a school across different grade levels fails to produce meaningful data for comparison purposes. The same obstacles that prevent comparison among standardized tests, as described by the Task Force, also prevent a meaningful aggregate of test scores per school. Standardized tests are not easily compared across grade levels within a school, and there is a lack of reliable data when comparing aggregate reports among a variety of standardized tests. Therefore, aggregate scores cannot be compared across schools.

SEAA respectfully requests that the requirement that schools with over 25 scholarship recipients submit an aggregate test score, as set forth in G.S. 115C-562.5(c), be removed. The term “aggregate test score” is not defined in the Program statute. If the term aggregate test score is intended to capture nonpublic schools with more than 25 scholarship students’ test scores aggregated or averaged together regardless of grade level, we note that schools have reported challenges in obtaining that information from testing companies. The primary reason is that aggregates are not commonly performed across grade levels. Due to these issues, SEAA has explored various options to assist nonpublic schools with meeting the current statutory requirement to report aggregate test scores. However, because of the reasons set forth in this report, it is not feasible for nonpublic schools to comply with the existing aggregate reporting requirement. In addition, an evaluation design that includes sampling of nonpublic

⁸ In addition, approximately 2,100 ESA+ scholarship recipients will also be required to be tested for the 2022-23 school year.

⁹ In accordance with G.S. 115C-549, 550 and G.S. 115C-557, 558, each private school must administer a nationally standardized test or other nationally standardized equivalent measurement selected by the chief administrator to all students enrolled and regularly attending grades 3, 6, 9, and 11 each year.

school and public school students by the independent research organization would negate the need for nonpublic schools to collect and report this aggregate data directly to SEAA. However, if the statutory requirement for aggregate data is not removed due to its use in conjunction with the learning gains and losses report, SEAA requests, in the alternative, that the statute define what an “aggregate test score” means, such that schools would be able to fulfill this requirement.

Initial Timeline and Implementation Costs

Initial Timeline

Fiscal Year 2023-24	SEAA selects research organization and structures the study.
Academic Year 2024-25	First testing of students in Cohort 1.
Academic Year 2025-26	Second testing of students in Cohort 1.
December 1, 2026	First learning gains and losses report.

Costs. SEAA estimates the cost to contract with a research organization on an annual basis would require an appropriation of between \$300,000 to \$450,000 in recurring funds. The cost variance is based on the research organization’s recommendation of sample size. The cost of the test would also depend on the number of public and private school students participating in the test and which test is chosen. We would expect the research organization to make this recommendation as a part of its study.

Study of Opportunity Scholarship Student Evaluations

North Carolina State Education Assistance Authority

Due March 1, 2018

Submitted March 1, 2018

**STUDY OF OPPORTUNITY SCHOLARSHIP
STUDENT EVALUATIONS**
Section 10A.6 of Session Law 2017-57

INTRODUCTION

In Section 10A.6(a) of Session Law 2017-57, “Current Operations Appropriations Act of 2017,” the General Assembly required that the State Education Assistance Authority (Authority) establish a task force to study the evaluation of students receiving scholarship grants through the Opportunity Scholarship Grant Program pursuant to G.S. 115C-562.7(c). The Authority collaborated with the Department of Administration, Division of Nonpublic Education, and the Department of Public Instruction to complete this study.

In addition to representatives from the three collaborating organizations, Task Force members included representatives from nonpublic schools of varying sizes, organizations representing parental school choice (specifically Parents for Educational Freedom in North Carolina), organizations representing nonpublic schools, two independent research organizations, and the public schools, as set out in the legislation. A list of Task Force members appears at the end of this report.

The Task Force met four times (October 23, 2017; November 29, 2017; January 24, 2018; and February 21, 2018) to consider the challenges to and possible methodologies for comparing learning gains or losses between participants in the Opportunity Scholarship Program and selected North Carolina public school students.

TASK FORCE CHARGE

The task force shall study the most effective, valid, and reliable method of evaluating learning gains or losses of students receiving scholarship grants and comparing the learning gains or losses of those students to public school students with similar socioeconomic backgrounds, including the potential for adoption of a nationally normed common test for students participating in the evaluation. In doing so, the task force shall also consider the most reliable manner of establishing causal relationships to student performance outcomes while achieving minimal interference with the operation of the participating nonpublic and public schools, including limited sampling and other suitable research design methods. -- S.L. 2017-57, SECTION 10A.6.(b)

The following terms or phrases from the statutory language are defined at the end of this report: *effective, valid, reliable, comparing learning gains or losses, nationally normed test, causal relationship, sampling, and concordance*. The Task Force agreed to these explanations in order to begin discussion of the assignment with a common language. In addition to a definition of the term

**STUDY OF OPPORTUNITY SCHOLARSHIP
STUDENT EVALUATIONS**
Section 10A.6 of Session Law 2017-57

causal relationship, relevant challenges to establishing a causal relationship are explained. All terms except for *concordance*, which emerged during Task Force discussions, occur in the language of Section 10A.6 of Session Law 2017-57.

RECOMMENDATION

The Task Force wishes to emphasize to the legislature the commonly held desire among all Task Force members to offer significant positive and measurable educational opportunities for students.

The Task Force has determined that meeting the statutory charge to provide “the most effective, valid, and reliable method” of evaluation and “the most reliable manner of establishing causal relationships to student performance outcomes” requires the following:

1. Evaluators select a random sample of students enrolled in a nonpublic school receiving the Opportunity Scholarship and a random sample of eligible students who applied for the Opportunity Scholarship but were not awarded funds and remained in public school. This second group of students is a comparison group.
2. Both groups of students take the same nationally normed test. Students in the study sample take the test a minimum of two times: once during the Opportunity Scholarship application process to establish baseline performance and again after one year to assess academic growth. To track trends over time in student achievement, additional test administrations (i.e., annually after the initial two test administrations) would be necessary.
3. Once Condition 1 and Condition 2 are met, researchers then will be able to compare the test score performance of students receiving Opportunity Scholarship funds to students in the comparison group and assess the causal impact of the Opportunity Scholarship Program.

This evaluation design is the *most* effective, valid, and reliable design possible, given the context, and would provide the *most* valid and reliable method of measuring a potential causal relationship between students’ use of the Opportunity Scholarship and their performance outcomes.

There is one major impediment, as well as other important considerations.

To date, the applications for an Opportunity Scholarship have not yielded sufficient numbers of qualifying non-awarded students to produce a comparison group (i.e., more applicants than

**STUDY OF OPPORTUNITY SCHOLARSHIP
STUDENT EVALUATIONS**
Section 10A.6 of Session Law 2017-57

available scholarships). Most of the eligible applicants in grades 2 and higher were offered scholarships.

Other considerations include:

- The public and nonpublic school communities discourage a mandate to administer an additional required test to students that is different from currently administered tests.
- Both the public and the nonpublic schools expressed the need for a different type of evaluation that captures the broader value of the school learning experience that cannot be measured by a nationally normed test alone. A richer approach to an evaluation, including qualitative measures, may be more useful for stakeholders and of interest to policymakers, although potentially more expensive. The Task Force determined that development of these other recommendations was outside the scope of its assignment.
- Costs of tests, third-party test administration, and operationalizing a causal study in an educational environment are significant. An appropriation should be included in any potential legislation to prevent a financial burden on participating schools.
- Administration of tests will incur modest disruption to both public and nonpublic schools for random samples of students.
- The nonpublic schools note that individual assessments used by their schools and submitted to the Authority are carefully selected to be the best measure of the schools' course of study. An imposed assessment may not accurately measure learning gains of students if that test is not aligned with the course of study.

In conclusion, by stipulating that the Task Force shall “consider the most reliable manner of establishing *causal relationships* to student performance outcomes,” the legislature has established a high bar for the evaluation. Other less rigorous evaluations may be attainable, albeit with the caveat that those outcomes – though still informative – will be less rigorous as well.

**STUDY OF OPPORTUNITY SCHOLARSHIP
STUDENT EVALUATIONS**
Section 10A.6 of Session Law 2017-57

DEFINITION OF TERMS

1. **Effective:** An effective evaluation method is one that fulfills its specified function. Note that *effective* in this case refers to the evaluation method, not to the program or to schools.
2. **Valid:** A valid evaluation method supports the types of inferences the researchers hope to draw. A valid evaluation means that the researchers are collecting the evidence needed to support the inferences or conclusions they wish to make from test scores.
3. **Reliable:** A reliable evaluation method is one that will consistently measure what the method intends to measure.
4. **Comparing learning gains or losses:** A research design must compare test scores from at least two time points to determine the magnitude and direction of any changes in student outcomes.
5. **Nationally normed test:** A nationally normed test (such as IOWA Test of Basic Skills ITBS or Terra Nova) is designed to focus on the comparison of a student's performance to that of a nationwide group of students who completed the same instrument. In contrast, a criterion-referenced test (such as the North Carolina public school End-of-Grade and End-of-Course tests) is designed to measure mastery of specific content knowledge and is not designed to allow a researcher to draw comparisons with other students.
6. **Causal relationship:** To identify a causal relationship, researchers must estimate how people who participated in a program fared compared to how they would have fared if they had not participated in the program. This hypothetical condition is called the counterfactual. Although the counterfactual is never directly observed (because an individual either experienced the program at a particular point in time or they did not), the research design needs to approximate the counterfactual in order to test whether or not a causal relationship exists.

Two challenges for researchers regarding establishing a causal relationship:

- a. Many factors other than the program can influence outcomes over time. Researchers cannot simply measure outcomes before and after the program. Therefore, researchers must track outcomes for program participants and for a comparison group.
- b. Based on differences in motivation, availability of information, and other factors, people who sign up for a program are systematically different from those who do not sign up for it. As a result, researchers must guard against what is referred to as "selection bias." Selection bias occurs when the researcher attributes differences in outcomes (positive or negative) to the program being evaluated, when those outcomes are, in fact, attributable to pre-existing differences in people who self-selected to participate in the program and those who did not. Therefore, researchers must identify treatment and comparison groups that, on average, have similar observed and unobserved characteristics.

**STUDY OF OPPORTUNITY SCHOLARSHIP
STUDENT EVALUATIONS**
Section 10A.6 of Session Law 2017-57

7. **Sampling:** A sample is a representative subset of a whole population selected for the purposes of drawing inferences about the population as a whole.
8. **Concordance:** The degree of agreement between two measures, such as two different tests, is known as concordance. Reliable test concordance can be established only when the same group of participants take the two tests that are being compared, or when a sufficiently large sample of students takes each test. In the absence of a well-designed concordance study, there may not be sufficient power to detect a statistically significant difference even if one exists.

TASK FORCE MEMBERSHIP AND AFFILIATION

Mr. Don Adams	Berean Baptist Academy
Ms. Diane Allen	Division of Nonpublic Education
Dr. Anna Egalite	North Carolina State University
Ms. Allison Guenther	Parents for Educational Freedom in North Carolina
Dr. Joe Haas	North Carolina Christian School Association
Dr. Rita Haire	Parents for Educational Freedom in North Carolina
Dr. Laura Knapp	RTI International
Ms. Linda Nelson	North Carolina Association of Independent Schools
Dr. Kathryn Marker	State Education Assistance Authority
Ms. Elizabeth McDuffie	State Education Assistance Authority
Dr. Bradley McMillen	Wake County Public School System
Dr. Perry Nicklow	Wesleyan Christian Academy
Dr. John Pendergrass	Catholic Diocese of Raleigh
Mrs. Cheryl Riley	Victory Christian Center School
Dr. Trip Stallings	North Carolina State University
Dr. Tom Tomberlin	Department of Public Instruction