THE GOVERNOR'S OFFICE OF STUDENT ACHIEVEMENT

Assessment Audit

February 21, 2019

Process Overview

The Governor's Office of Student Achievement (GOSA) serves as the reporting and accountability agency for education in Georgia. As such, GOSA is charged by law (O.C.G.A. § 20-14-35 and O.C.G.A. § 20-14-36) with inspecting academic records of schools to ensure that education institutions are faithful to performance accountability requirements. Through an academic audit, GOSA reviews student assessment data and other school records reported to the Georgia Department of Education (GaDOE) to confirm accuracy and explore the effectiveness of local school initiatives in improving achievement.

Data from state standardized assessments aim to assist in making educational policy decisions and provide a measure of students' academic performance, as well as the schools' effectiveness and adherence to prescribed standards. The Assessment Division at GaDOE oversees the development and administration of the Georgia Milestones End of Grade (EOG) for grades 3-8 and Georgia Milestones End of Course (EOC) assessments in ten high school courses.

GOSA conducts an annual assessment audit to ensure that all schools and local education agencies (LEAs) administer assessments with fidelity. Through the assessment audit, GOSA also reviews details on the administration of the Georgia Alternate Assessment 2.0 (GAA 2.0), as well as ACCESS and Alternate ACCESS for English Language Learners 2.0 (ACCESS for ELLs 2.0) for audited schools.

Phase 1: Data Review

The Georgia Milestones testing vendor, Data Recognition Corporation (DRC), is responsible for scoring the Georgia Milestones exams and reporting the results to GaDOE and the LEAs. Given the importance of these assessments, GOSA, as part of its statutory role, partners with DRC to conduct a comprehensive examination of all statewide answer documents for all EOG and EOC assessments.

The DRC analysis includes the following assessments:

- EOGs (*Spring 2018 administration only*)
    - English/Language Arts (grades 3-8)
    - Mathematics (grades 3-8)
    - Science (grades 5 and 8)
    - Social Studies (grades 5 and 8)
- EOCs (*Winter 2017, Spring 2018, and Summer 2018 administrations*)
    - Algebra I
    - American Literature and Composition
    - Analytic Geometry
    - Coordinate Algebra
    - Biology

- Economics
- Geometry
- Ninth Grade Literature and Composition
- Physical Science
- United States History

For the 2017-2018 school year, DRC reported analyses of wrong-to-right answer changes, response times, gain scores, and unusual response patterns. These analyses identify classrooms and schools for which these metrics are well above the state average. GOSA uses the four analyses to flag schools for the assessment audit (see appendix for more calculation details). All calculations exclude classrooms with fewer than eight students. It is important to note that the results of these analyses are used as an initial flag to spur further investigation of many indicators to determine if any cheating occurred. The results of the analyses do not indicate that cheating necessarily occurred.

Answer Change Analysis Flag[1]

- Five percent or more of classrooms in a school have a wrong-to-right standard score greater than or equal to 4.0 for EOGs OR greater than or equal to 5.0 for EOCs,[2] *OR*
- One classroom in a school has a wrong-to-right standard score greater than or equal to 7.0 for EOGs or EOCs.

Unusual Response Pattern Analysis Flag

- Two or more testing groups in a school are in the 95th percentile or greater for unusual responses.[3]

Gain Score Analysis Flag

- One or more classrooms have an outlier score of 10 or greater.

Response Time Analysis Flag

- Thirty percent or more of classrooms in a school have a test duration standard score greater than or equal to 4.0 or less than or equal to -4.0 for EOGs OR greater than or equal to 5.0 or less than or equal to -5.0 for EOCs, *OR*
- One classroom in a school has a test duration standard score greater than or equal to 7.0 or less than or equal to -7.0 for EOGs or EOCs.

[1] The Answer Change Analysis was known as the Erasure Analysis prior to the 2016 Georgia Milestones. Beginning in 2016, the Answer Change Analysis included paper-and-pencil and online administrations.

[2] The Answer Change Analysis and Response Time Analysis use a standard score (z-score) to control standard deviation for differences in classroom size.

[3] The Unusual Response Pattern analysis uses the calculation methodology of Jacob and Levitt (2003) and only includes assessments in English/Language Arts and Mathematics, starting with 4th grade. Each testing group is the total number of students by grade level and subject area (ELA or mathematics) who took a certain test form (A or B) regardless of classroom assignment. For example, all students in a school who took the 4th grade mathematics Georgia Milestones Form A assessment are a testing group.

Phase 2: Notification of Audit and Inquiry Form

GOSA presents the findings of the DRC analyses, along with recommendations, annually to the State Board of Education (SBOE). These recommendations, which the SBOE votes to approve, include requiring LEAs to conduct internal investigations to determine the causes of testing irregularities via an inquiry form. GOSA may also recommend that schools rotate teachers during test administration, so that they administer the test to students they have not taught. In addition, GOSA may place state monitors in these flagged schools during the Spring 2019 Georgia Milestones administration.

Schools may receive a flag for each analysis for a maximum of four flags. GOSA will flag any school receiving two or more flags for the assessment audit. GOSA will notify LEAs of any flagged schools and request all flagged schools to complete an inquiry form.

The school's testing coordinator will complete the inquiry form providing details about the school's 2017-2018 Georgia Milestones, 2018-2019 GAA 2.0, and 2018-2019 ACCESS and Alternate ACCESS for ELLs 2.0 administrations. Testing coordinators will also provide the testing plan and logistics for the Spring 2019 Georgia Milestones administration in preparation for a potential test monitoring visit.

GOSA will also send the inquiry form to schools who require further attention based on the results of previous audit years. State charter schools that opened in the last two academic years are included in the assessment audit and will also complete the inquiry form, excluding the portion related to the 2017-2018 Georgia Milestones administration.

Phase 3: Inquiry Form Review and Test Monitoring Determination

Due to limited staff capacity, GOSA will monitor about two percent of the total number of schools with Milestones data. GOSA will also monitor state charter schools which opened in the last two academic years as these schools establish their testing practices. State charter schools and schools requiring continued monitoring based on prior assessment audits are included in the two percent of schools for test monitoring.

After accounting for state charter schools and schools identified by prior assessment audits, GOSA selects schools to test monitor from the flagged schools that submitted inquiry forms (schools with two or more flags). To determine schools for test monitoring, GOSA considers the following:

- Total number of flags;
- 2017-2018 Georgia Milestones administration information provided in the inquiry form; and
- Random selection.

GOSA will not notify selected schools of the test monitoring visit; <u>any school completing the inquiry form should expect a potential test monitoring visit</u>.

Phase 4: Test Monitoring Visit

GOSA sends a staff member to each school selected for test monitoring to observe one day of Georgia Milestones test administration. The GOSA staff member will observe test security practices and fidelity to the testing instructions described in GaDOE's assessment manuals, including instructions for tests with accommodations for students with disabilities and English language learners.

Phase 5: Reporting and Closeout

GOSA reviews the test monitoring forms and determines appropriate next steps. If GOSA requires no further inquiry for the LEA, then GOSA notifies the superintendent. For schools that require further investigation, either the LEA or GOSA will refer the case to the Georgia Professional Standards Commission (GaPSC). GaPSC is statutorily responsible for regulating professional employees in Georgia's public schools by investigating allegations of educator misconduct and providing recommendations for disciplinary actions.

References

Jacob, B. & Levitt, S (2003). Rotten apples: An Investigation of the prevalence and predictors of teacher cheating, *The Quarterly Journal of Economics*, 118 (3), 843-877.

Appendix: Flagging Criteria Calculation Guide

Data Recognition Corporation (DRC) processes data from the Georgia Milestones administration and reports various metrics to the Georgia Department of Education (GaDOE). The Governor's Office of Student Achievement (GOSA) uses these metrics to determine which schools to include in the assessment audit. The guide below outlines the flagging criteria GOSA uses to select schools for the assessment audit and includes an overview of each calculation.

None of the criteria listed indicate whether a classroom or school engaged in unethical testing practices. Each is simply an indicator of behaviors outside of the state norm that might indicate unusual practices.

DRC only calculated the Unusual Response Pattern and Gain Score flags for classrooms with eight or more students. GOSA further excluded classrooms with fewer than eight students from the other two analyses.

Answer Change Analysis

Calculation Overview
The Answer Change Analysis looks at instances in which students changed their answer choice from a wrong answer to a right answer. For online administrations, the INSIGHT test administration software captures answer changes during online testing sessions. For paper-and-pencil administrations, optical scanning software reviews the integrity of each answer document according to pre-defined guidelines.

The Answer Change Analysis uses standard scores (z-scores, instead of strictly standard deviation) for the Answer Change Analysis to account for differences in classroom sizes:

$$z = \frac{x - \mu}{\sigma/\sqrt{n}}$$

x = classroom mean of wrong-to-right answer changes
μ = mean wrong-to-right of state population
σ = standard deviation of state population
n = classroom size

Flagging Criteria
GOSA flags schools through the Answer Change Analysis if the school meets at least one of the below criteria:

- Five percent or more of classrooms in a school have a wrong-to-right standard score greater than or equal to 4.0 for EOGs OR greater than or equal to 5.0 for EOCs. This criterion identifies schools in which several classrooms demonstrate incidences of wrong-to-right answer changes that are higher than the state norm.
- One or more classrooms in a school has a wrong-to-right standard score greater than or equal to 7.0 for EOGs or EOCs. This criterion identifies classrooms with incidences of wrong-to-right answer changes that are substantially higher than the state norm.

Unusual Response Pattern Analysis

Calculation Overview
DRC reports the Unusual Response Pattern analysis for English/Language Arts and Mathematics assessments beginning with 4th grade. The calculation is based on the methodology of Jacob and Levitt (2003).[4] The Unusual Response Pattern analysis includes two indices. The first index ranks each classrooms' average test score gains relative to other classrooms in that grade and subject, based on matched student test data. The ranking is scaled to percentiles.

The second index includes unexpected patterns in student answers, considered four ways:

- Unlikely blocks of identical answers given by students on consecutive items,
- Degree of correlation in student responses across the test (particularly for unexpected answers),
- Variance of test residuals for the cohort (particular questions having extremely high residual deviations within cohorts), and
- Cases in which students miss easy items while answering difficult items correctly.

DRC then ranks classrooms across each of these four measures. The ranks are combined to create a composite index ranking, scaled to percentiles.

Flagging Criteria
GOSA flags schools for unusual response patterns if the school meets the below criterion:

- Two or more testing groups in a school were in the 95th percentile or greater for unusual responses, meaning the testing group ranked in the 95th percentile for both indices.

Gain Score Analysis

Calculation Overview
DRC reports the Gain Score Analysis for English/Language Arts and Mathematics assessments beginning with 4th grade. Using the classroom administrator of record in 2018, the difference in mean scale scores for the group of students associated with the administrator is compared against the difference in scale scores for the state.[5] Students are matched from 2017 to 2018. Gain score calculations are based on t-test results between gains in the testing group versus the state.

The outlier score is calculated using the p-value of the t-test:

$$\text{OS} = \left|1.0861 ln\left(\frac{p}{q}\right)\right|$$

p = p-value of the t-test, or the probability that the classroom gains are significantly higher than the average state gains
$q = 1 - p$

The coefficient 1.0861 was used such that a p-value of .0001 (significantly low probability that such gains would occur) results in an outlier score of 10. The natural log was taken to make the scale symmetric around small and large probabilities. Outlier scores are reported on a scale of 0 to 50. An outlier score of 10 or greater is considered different from the baseline.

[4] Jacob, B. & Levitt, S (2003). Rotten apples: An Investigation of the prevalence and predictors of teacher cheating, *The Quarterly Journal of Economics*, 118 (3), 843-877.

[5] DRC converts the scaled scores to z-scores before computing the statistical test to account for differences in scale scores across grades.

Flagging Criteria
GOSA flags schools for gain scores if the school meets the below criterion:

- One or more classrooms has an outlier score of 10 or greater. This means the probability that a classroom's gains are equal to the average state gains is less than 0.01%.

Response Time Analysis

Calculation Overview
DRC reports the Response Time Analysis for online administrations of the Georgia Milestones. The INSIGHT test administration software captures test duration for all operational test items for each student.

The Response Time Analysis uses standard scores (z-scores, instead of strictly standard deviation) to account for differences in classroom sizes:

$$z = \frac{x - \mu}{\sigma/\sqrt{n}}$$

x = classroom mean test duration
μ = mean test duration of state population
σ = standard deviation of state population
n = classroom size

Flagging Criteria
GOSA flags schools for response times if the school meets at least one of the below criteria:

- Thirty percent or more of classrooms in a school have a test duration standard score greater than or equal to 4.0 or less than or equal to -4.0 for EOGs OR greater than or equal to 5.0 or less than or equal to -5.0 for EOCs. This criterion identifies schools in which several classrooms administered tests for a time period that was lower or higher than the state norm.
- One or more classrooms in a school has a test duration standard score greater than or equal to 7.0 or less than or equal to -7.0 for EOGs or EOCs. This criterion identifies classrooms that administered tests for a time period substantially higher or lower than the state norm.

Multiple Administrations

DRC analyzes data from the Spring EOG administrations and Winter, Spring, and Summer EOC administrations. GOSA calculates the four flags separately for each administration and considers each school's administration with the maximum number of flags when selecting schools for the audit. For schools that administer the Spring EOG and the Spring EOC (mostly middle schools or schools serving students in those grades), their Spring EOG and Spring EOC flags are combined, with a maximum of one flag per analysis.

Example for schools administering Spring EOG and Spring EOC:

	Answer Change	Response Pattern	Gain Score	Response Time	Number of Flags
Spring EOG		⚑		⚑	2
Spring EOC	⚑	⚑			2
School's Flags	⚑	⚑		⚑	**3**