# Spring 2015 Georgia Milestones EOG/EOC Erasure Analysis Desktop Audit Results

*February 2016*

THE GOVERNOR'S OFFICE OF
**STUDENT ACHIEVEMENT**

## Executive Summary

The Governor's Office of Student Achievement serves as the reporting and accountability agency for education in Georgia. As such, GOSA is charged by law 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 State to confirm accuracy and explore the effectiveness of local school initiatives in improving achievement.

The erasure analysis focuses on identifying classrooms and schools where the number of wrong answers that have been changed to right answers on individual student answer sheets is well above the state average. It is conducted in English-Language Arts, Mathematics, Science, and Social Studies in grades 3 through 8 and following eight high school courses: Ninth Grade Literature and Composition, American Literature and Composition, Coordinate Algebra, Analytic Geometry, Physical Science, Biology, U.S. History, and Economics. It is important to note that the results of the erasure analysis are used as an initial flag to spur further investigation of many indicators to determine if any cheating occurred. The results do not indicate that cheating necessarily occurred.

To conduct this analysis, the State's testing vendor, Data Recognition Corporation (DRC) had psychometricians scan answer documents to identify total erasures and wrong-to-right erasures per classroom.[1] Using the DRC Erasure Analysis and accompanying data file, GOSA flagged schools for an internal desktop audit based on the following criteria: [2]

**EOG Flagging Criteria (Grades 3-8)**

- Five percent or more of classrooms in a school are flagged at four standard deviations or greater, OR
- One classroom is flagged at seven standard deviations or greater.

**EOC Flagging Criterion (Grades 7-12)**

- One classroom is flagged at five standard deviations or greater.

**Results Summary**

For the EOG, 94 classrooms in 40 schools in 23 LEAs were identified for an initial desktop audit. After the audit, 19 schools in 12 LEAs require further inquiry. For the EOC, 32 classrooms in 23 schools in 12 LEAs were identified for an initial desktop audit. After the audit, 14 schools in 10 LEAs require further inquiry.

---

[1] Currently, the analysis only includes paper-and-pencil tests. As part of its FY16 contract, DRC is developing pilot analysis of wrong-to-right answer changes from online testing. In addition, due to the technology in online testing, it is also providing pilot response similarity and response time analyses that may strengthen GOSA's auditing efforts in future years.

[2] In prior years, GOSA placed schools into four categories based on the percentage of classrooms flagged within each school: Clear of concern; Minimal concern; Moderate concern; and Severe concern. As a result, schools with a classroom flagged with a high standard deviation were not identified for a desktop audit if less than 5% of the classrooms in a school were flagged. With this in mind, GOSA has adjusted the standard deviation levels for EOG and EOC and added a criterion that automatically flags classrooms with greater than 7 standard deviations.

Schools requiring further inquiry are included in recommendations to the SBOE for inquiry, monitoring, teacher rotation during testing, and on-site audits. The following report contains the results of GOSA's desktop audit and recommendations to the State Board of Education for actions to be taken in schools requiring further inquiry by the State.

## Table of Contents

## Table of Abbreviations

| GaDOE | Georgia Department of Education | LEA | Local Education Agency (District) |
|-------|-------------------------------|------|----------------------------------|
| EOG | End of Grade Test | GOSA | Governor's Office of Student Achievement |
| EOC | End of Course Test | SBOE | State Board of Education |
| SD | Standard Deviation | | |

## EOG Desktop Audit

**Analysis Overview**

GOSA identified schools for a desktop audit when five percent or more of classrooms in a school was flagged at four standard deviations or greater, or one classroom was flagged at seven standard deviations or greater. In total, 94 classrooms in 40 schools in 23 LEAs were identified for an initial desktop audit to determine a possible explanation for the flag that would remove the need for further inquiry.

In the desktop audit analysis, many school-level factors, outlined in detail in Appendix A, were reviewed holistically and discussed as a team before any determinations were made. GOSA placed schools in one of two categories: "further inquiry needed," or "no further inquiry needed."

After the desktop audit, 19 schools in 12 LEAs, listed below, require further inquiry. Appendix B lists the number of classrooms at each school requiring further inquiry for both EOG and EOC.

**Desktop Audit Schools**

*Schools Requiring Further Inquiry*

- Scott Elementary School (Atlanta Public Schools)
- Miles Intermediate School (Atlanta Public Schools)
- Cleveland Elementary School (Atlanta Public Schools)
- Sequoyah Middle School (DeKalb)
- Stephenson Middle School (DeKalb)
- Stone Mountain Middle School (DeKalb)
- Montgomery Elementary School (DeKalb)
- Huddleston Elementary School (Fayette)
- Taylor Road Middle School (Fulton)
- Fulton Sunshine Academy Charter School (Fulton)
- Risley Middle School (Glynn)
- Greensboro Elementary School (Greene)
- Harmony Elementary School (Gwinnett)
- Friendship Elementary School (Hall)
- Langston Road Elementary (Houston)
- Blackburn Elementary School (Lumpkin)
- Rosemont Elementary School (Troup)

## EOC Desktop Audit

Schools were identified for a desktop audit when one classroom was flagged at five standard deviations or greater within the school. Thirty-two classrooms in 23 schools in 12 LEAs were identified for an initial desktop audit to determine a possible explanation for the flag that would remove the need for further inquiry.

In this analysis, many school-level factors were reviewed holistically and discussed as a team before any determinations were made. GOSA placed schools in one of two categories: "further inquiry needed," or "no further inquiry needed."

After the desktop audit, 14 schools in 10 LEAs, listed below, require further inquiry. Appendix B lists the number of classrooms at each school requiring further inquiry for both EOG and EOC.

**EOC Desktop Audit Schools**

*Schools Requiring Further Inquiry*

- Early College High School at Carver (Atlanta Public Schools)
- Charles R Drew High School (Clayton)
- Sprayberry High School (Cobb)
- Clarkston High School (DeKalb)
- Peachtree Middle School (DeKalb)
- Tucker High School (DeKalb)
- Tucker Middle School (DeKalb)
- Chapel Hill High School (Douglas)
- Alpharetta High School (Fulton)
- North Springs High School (Fulton)
- West Hall High School (Hall)
- Jefferson High School (Jefferson City)
- Jordan Vocational High School (Muscogee)
- Southeast Whitfield High School (Whitfield)

## GOSA Recommendations in Schools Requiring Further Inquiry

Overall, 85 classrooms in 33 schools in 18 LEAs require further inquiry. GOSA recommends the following actions to the SBOE in these schools:

- State monitors will observe and inspect schools requiring further inquiry for the 2016 Georgia Milestones test administration (EOG and EOC tests).
- Schools must rotate teachers in schools requiring further inquiry for the 2016 Georgia Milestones (EOG tests).
- GOSA will share data files with superintendents to facilitate:
    - LEA investigation of reason(s) for flags
    - Submission of online inquiry form to GOSA with results of investigation and an explanation of testing protocols in place.
- GOSA will conduct on-site audits as necessary.

No further action should be taken for flagged schools that require no further inquiry after the desktop audit.

## Appendix A: Desktop Audit Indicators

| Desktop Audit Indicators Reviewed |
|---|
| Number of classrooms flagged in each school and whether the flagged classrooms had different test administrators. |
| Total erasures and number of wrong-to-right (w-t-r) at the classroom level, including student-level data to determine whether erasures are concentrated in a small number of students. Classrooms where more than 50% of students in a classroom have zero erasures and/or w-t-r erasures reduce the likelihood of systematic or widespread changes in answers from wrong to right. |
| The severity of the individual flagged classroom (i.e. the standard deviation value or how far from what is considered normal behavior is the class positioned). EOC flags between 5.0 and 6.0 SDs are of less concern than those over 6.0 SDs. |
| Percentage of total classroom erasures changed from w-t-r. Generally, classrooms with greater than 65% of erasures being w-t-r are of concern, unless a classroom with multiple students had one student with many w-t-r erasures, suggesting that systematic cheating was unlikely. |
| The number of students in each classroom. (Example: Extremes in classroom populations on both ends of the distribution can skew post-calculation metrics and in turn cause flagged classrooms.). |
| Classroom percentile ranks of wrong-to-right erasures by student to observe the distribution of erasures in a classroom and compare that distribution to the state distribution. For example, comparing a classroom's 50th and 90th percentile with the state 50th and 90th percentiles can identify whether abnormal distributions and/or outliers. |
| The type of school (i.e. state charter school, high transient population, alternative education program, residential treatment facilities, etc.). |
| School demographics and groups (ELL population, gifted, magnet, students with disabilities, etc.). |
| Variance in performance level data from previous years (not applicable in 2014-2015 due to Georgia Milestones transition). |
| History as a school of concern. |
| Prior test monitoring and/or an on-site audit by state personnel. |
| District personnel and/or policies currently implemented to support test security. |
| Review of state monitor notes and/or forms. |

## Appendix B:  EOG/EOC Schools Requiring Further Inquiry

The following list includes the number of classrooms flagged in the 33 schools requiring further inquiry after the desktop audit.

| System Name | School Name | 2015 Spring EOG Classes Requiring Further Inquiry | 2015 Spring EOC Classes Requiring Further Inquiry |
|---|---|---|---|
| Atlanta Public Schools | Scott Elementary School | 4 | |
| Atlanta Public Schools | Miles Intermediate Elementary School | 5 | |
| Atlanta Public Schools | Cleveland Elementary School | 2 | |
| Atlanta Public Schools | Early College at Carver | | 1 |
| Clayton County Schools | Charles R Drew High School | | 1 |
| Cobb County Schools | Sprayberry High School | | 1 |
| DeKalb County Schools | Stone Mountain Middle School | 21 | |
| DeKalb County Schools | Montgomery Elementary School | 3 | |
| DeKalb County Schools | Sequoyah Middle School | 2 | |
| DeKalb County Schools | Stephenson Middle School | 1 | |
| DeKalb County Schools | Clarkston High School | | 9 |
| DeKalb County Schools | Peachtree Middle School | | 1 |
| DeKalb County Schools | Tucker High School | | 1 |
| DeKalb County Schools | Tucker Middle School | | 1 |
| Douglas County Schools | Chapel Hill High School | | 1 |
| Fayette County Schools | Huddleston Elementary School | 3 | |
| Forsyth County Schools | Brookwood Elementary School | 1 | |
| Fulton County Schools | Fulton Sunshine Academy | 2 | |
| Fulton County Schools | Taylor Road Middle School | 2 | |
| Fulton County Schools | Alpharetta High School | | 1 |
| Fulton County Schools | North Springs High School | | 1 |
| Glynn County Schools | Risley Middle School | 1 | |
| Greene County Schools | Greensboro Elementary School | 1 | |
| Gwinnett County Schools | Nesbit Elementary School | 2 | |
| Gwinnett County Schools | Harmony Elementary School | 2 | |
| Hall County Schools | Friendship Elementary School | 3 | |
| Hall County Schools | West Hall High School | | 1 |
| Houston County Schools | Langston Road Elementary School | 3 | |
| Jefferson City Schools | Jefferson High School | | 1 |
| Lumpkin County Schools | Blackburn Elementary School | 4 | |
| Muscogee County Schools | Jordan Vocational High School | | 1 |
| Troup County Schools | Rosemont Elementary School | 1 | |
| Whitfield County Schools | Southeast Whitfield High School | | 1 |
| TOTAL (18 LEAs) | TOTAL (33 schools) | 63 | 22 |

## Appendix C:  2015 DRC EOG Executive Report

**Erasure Analysis**

**Submitted by DRC**

**November 2015**

With the high-stakes nature of large-scale assessments such as the Milestones End of Grade (EOG), there are times when students' responses, and hence their scores, may not be a true representation of their own abilities. Various activities may take place, such as a student copying from another student's paper, students receiving inappropriate assistance before or during testing, or students' responses altered after testing. To maintain the integrity of the Milestones EOG and the validity of the results, it is important that any such instances be discovered.

The present study investigated student responses on the English Language/Arts, Mathematics, Science and Social Studies tests of the 2015 Spring Milestones EOG that a) were erased and b) changed from a wrong answer to a right answer (wrong-to-right).

It should be emphasized that results from the erasure analyses performed in 2015 should only be used to identify potential problems within individual classrooms. That is, these types of analyses must be supported by additional, collateral information before conclusions regarding any improprieties are reached.

### Scanning Operations

The GA Milestones EOG answer documents were processed using high speed 5000i optical scanners which reliably captured document images and optical mark read data. The sophisticated proprietary scoring software system, specifically Optical Mark Recognition (OMR) software, reviews the integrity of each batch of documents scanned according to pre-defined guidelines and services.

The OMR software provides a mechanism for identifying multiple-marks and identification of erasures for scanned data. The basis of the erasure analysis is to count erasures for multiple-choice items where two or more responses have been made with a specified intensity. Erasure analyses provide a mechanism to differentiate between three kinds of answer changes: a) wrong-to-wrong, b) right-to-wrong and c) wrong-to-right. Capturing the frequency of answer changes from wrong-to-right can be useful for identifying potential instances of cheating at the student level. Erasure analyses results can be grouped to tentatively identify problems at the classroom and school levels.

### Method

The basis for the erasure analysis is to count erasures in items where an answer choice was erased and replaced with another answer choice. Often the data captured is useful for identifying cases of cheating. During erasure analysis, two sets of erasures were analyzed: all erasures and wrong-to-right erasures where an incorrect answer choice was erased and replaced

1

with the correct answer choice. Please note that, for the erasure analyses, all items (both operational and field-test) were included, as all field-test items were embedded in the Milestones EOG.

The basic idea underlying the procedure is a statistical test of the null hypothesis ($H_0$) that the mean number of erasures for a class constitutes a random sample from the state distribution of erasures. The hypothesis is tested against the (right-sided) alternative ($H_1$) that the mean number is too high to be explained by random sampling. Classes for which $H_0$ has to be rejected are flagged for further scrutiny. A well-known central limit theorem in statistics tells us that the sampling distribution of the mean number of erasures for class $i$ ($m_i$) is asymptotically normal with mean and standard deviation (SD)

$$mean(m_i) = \mu \qquad (1)$$

$$SD(m_i) = \frac{\sigma}{\sqrt{n_i}} \qquad (2)$$

where $n_i$ and $m_i$ denote the size and mean number of erasures for class $i$, respectively. In addition, $\mu$ and $\sigma$ denote the mean and the SD of the distribution of the number of erasures of the population of individual students in the state of Georgia.

The classes were flagged if their $m_i$ was larger than $\mu + 4\frac{\sigma}{\sqrt{n_i}}$. Statistically, the flagging criterion set at or above $4\sigma$ is conservative. The standard normal table shows that under random sampling the (asymptotic) probability of a sample mean being more than four SDs above the population mean is around 0.00003. However, rejection of $H_0$ only tells us that the observed mean number of erasures is unlikely to be the result of random sampling.

It is evident in the formula that the class flagging criterion for each class is adjusted for the number of test takers in a classroom. For example, if the state mean and SD of erasure count are 1.73 and 2.11, respectively, the flagging criterion for a class size of 20 is adjusted to 3.62 ($1.73 + 4\frac{2.11}{\sqrt{20}} = 3.62$ ).

This adjustment ensures that the flagging criterion is equally stringent for classes with considerably different numbers of test takers. In addition, minimizing the probability of false positive (Type I) errors in this statistical test is crucial in this analysis.

<div align="center">

**Results**

</div>

Table 1 reports the state summary of erasure counts. The table includes the number of students, the total number of all types of erasures, the mean and the SD of all types of erasures, the correlation between all erasures and wrong-to-right erasures, the number of erasures at the 50th, 75th, 90th, 95th, 99th, and 99.9th percentiles, and the maximum number of all types of erasures. The mean number of erasures across all courses ranged from 1.39 to 3.41 for the 2015 Spring Milestones EOG. In other words, approximately 1 to 3 answer changes were made per student answer sheet on average. The erasure count at specific percentile points (50th, 75th, 90th,

2

95th, 99th, and 99.9[th]) is also reported. The erasure count at the 95th percentile point was between 5 and 10.

Table 2 reports the state summary of wrong-to-right erasure counts. The table includes the number of students, the number of wrong-to-right erasures, the mean and the SD of wrong-to-right erasures, the correlation between all erasures and wrong-to-right erasures, the number of wrong-to-right erasure at the 50th, 75th, 90th, 95th, 99th, and 99.9[th] percentiles, and the maximum number of wrong-to-right erasures. As can be expected, the mean wrong-to-right erasure count and the count at the specific percentile points were lower than those obtained from all erasure counts. The mean number of wrong-to-right erasures ranged from 0.75 to 2.04 for the 2015 Spring Milestones EOG. In other words, approximately 1 to 2 wrong-to-right answer changes were made per student answer sheet on average. The wrong-to-right erasure count at specific percentile points (50th, 75th, 90th, 95th, 99th, and 99.9[th]) is also reported. The wrong-to-right erasure count at the 95th percentile point was between 3 and 7.

Table 3 presents a summary of the number of schools flagged across four content areas - English Language/Arts, Mathematics, Science and Social Studies – within each analysis of the Milestones EOG. For each analysis, the number of schools was computed in two ways: flagged for at least one content area or flagged for all four content areas. The number/percentage of schools that had zero flags for all erasures and wrong-to-right erasures in English Language/Arts, Mathematics, Science and Social Studies is provided in Table 4. The number/percentage of schools that had less than 1% of the classes flagged for all erasures and wrong-to-right erasures in English Language/Arts, Mathematics, Science and Social Studies and across grades is provided in Table 5.

### Discussion

With respect to the erasure analyses, the following caveats are always applicable:

1. The normal distribution holds only for large classes; for smaller classes the result is approximate.

2. Rejection of $H_0$ does not necessarily imply cheating. Alternative explanations are possible.

3. The flagging criterion should thus be taken as a stimulus to look for additional evidence and find out what happened in the school.

This erasure analysis is considered a check for unusual numbers of erasures to student responses. Without additional layers added to the analysis, this kind of check only addresses the possibility, not the certainty, of teachers or administrators altering the responses of students. The 2015 erasure analyses represent an important step in helping to maintain the integrity of future administrations of the Milestones EOG.

3

**Table 1. State Summary Statistics for All Types of Erasure (ERA) Counts by Content/Grade**

| Content | Grade | N | No. of Erasures | Mean | SD | Correlation between ERA and WTR | Number of Erasure by Percentiles | | | | | | Max |
|---------|-------|---|-----------------|------|----|-----|----|----|----|----|----|------|-----|
| | | | | | | | 50 | 75 | 90 | 95 | 99 | 99.9 | |
| ELA | 3 | 106900 | 210773 | 1.97 | 2.18 | 0.77 | 1 | 3 | 5 | 6 | 10 | 17 | 36 |
| | 4 | 78078 | 122679 | 1.57 | 1.94 | 0.81 | 1 | 2 | 4 | 5 | 8 | 14 | 55 |
| | 5 | 89365 | 134836 | 1.51 | 1.85 | 0.79 | 1 | 2 | 4 | 5 | 8 | 15 | 32 |
| | 6 | 91830 | 127422 | 1.39 | 1.75 | 0.80 | 1 | 2 | 3 | 5 | 8 | 14 | 48 |
| | 7 | 86432 | 126594 | 1.46 | 1.86 | 0.80 | 1 | 2 | 4 | 5 | 8 | 14 | 29 |
| | 8 | 91202 | 140932 | 1.55 | 1.86 | 0.83 | 1 | 2 | 4 | 5 | 8 | 15 | 27 |
| MA | 3 | 107415 | 247973 | 2.31 | 2.44 | 0.81 | 2 | 3 | 5 | 7 | 11 | 17 | 67 |
| | 4 | 78603 | 137586 | 1.75 | 2.02 | 0.79 | 1 | 3 | 4 | 6 | 9 | 15 | 63 |
| | 5 | 89923 | 171436 | 1.91 | 2.15 | 0.81 | 1 | 3 | 5 | 6 | 9 | 16 | 32 |
| | 6 | 92145 | 159945 | 1.74 | 2.00 | 0.80 | 1 | 3 | 4 | 5 | 9 | 15 | 43 |
| | 7 | 86701 | 156588 | 1.81 | 2.06 | 0.79 | 1 | 3 | 4 | 6 | 9 | 15 | 32 |
| | 8 | 91500 | 177411 | 1.94 | 2.17 | 0.79 | 1 | 3 | 5 | 6 | 10 | 17 | 46 |
| SC | 3 | 107293 | 303908 | 2.83 | 3.15 | 0.83 | 2 | 4 | 7 | 9 | 14 | 27 | 62 |
| | 4 | 78507 | 157155 | 2.00 | 2.45 | 0.84 | 1 | 3 | 5 | 7 | 11 | 19 | 68 |
| | 5 | 89823 | 178940 | 1.99 | 2.37 | 0.84 | 1 | 3 | 5 | 6 | 11 | 18 | 50 |
| | 6 | 92034 | 179062 | 1.95 | 2.29 | 0.82 | 1 | 3 | 5 | 6 | 10 | 18 | 55 |
| | 7 | 86556 | 149609 | 1.73 | 2.15 | 0.82 | 1 | 2 | 4 | 6 | 10 | 17 | 36 |
| | 8 | 91347 | 164961 | 1.81 | 2.18 | 0.81 | 1 | 3 | 4 | 6 | 10 | 18 | 59 |
| SS | 3 | 106857 | 364170 | 3.41 | 4.13 | 0.89 | 2 | 5 | 7 | 10 | 18 | 49 | 69 |
| | 4 | 78167 | 142632 | 1.82 | 2.51 | 0.86 | 1 | 3 | 5 | 6 | 11 | 22 | 71 |
| | 5 | 89503 | 188224 | 2.10 | 2.64 | 0.86 | 1 | 3 | 5 | 7 | 11 | 23 | 59 |
| | 6 | 91744 | 168883 | 1.84 | 2.59 | 0.85 | 1 | 3 | 4 | 6 | 11 | 25 | 64 |
| | 7 | 86183 | 154059 | 1.79 | 2.58 | 0.87 | 1 | 2 | 4 | 6 | 10 | 26 | 67 |
| | 8 | 90927 | 175324 | 1.93 | 2.72 | 0.86 | 1 | 3 | 5 | 6 | 11 | 32 | 68 |

4

**Table 2. State Summary Statistics for Wrong-to-Right (WTR) Erasure Counts by Content/Grade**

| Content | Grade | N | No. of Erasures | Mean | SD | Correlation between ERA and WTR | Number of Erasure by Percentiles | | | | | | Max |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 50 | 75 | 90 | 95 | 99 | 99.9 | |
| ELA | 3 | 106900 | 102270 | 0.96 | 1.27 | 0.77 | 1 | 1 | 3 | 3 | 5 | 9 | 16 |
| | 4 | 78078 | 67613 | 0.87 | 1.23 | 0.81 | 0 | 1 | 2 | 3 | 5 | 9 | 41 |
| | 5 | 89365 | 67756 | 0.76 | 1.11 | 0.79 | 0 | 1 | 2 | 3 | 5 | 8 | 22 |
| | 6 | 91830 | 69227 | 0.75 | 1.10 | 0.80 | 0 | 1 | 2 | 3 | 5 | 8 | 25 |
| | 7 | 86432 | 65001 | 0.75 | 1.14 | 0.80 | 0 | 1 | 2 | 3 | 5 | 8 | 23 |
| | 8 | 91202 | 81255 | 0.89 | 1.22 | 0.83 | 1 | 1 | 2 | 3 | 5 | 9 | 16 |
| MA | 3 | 107415 | 142167 | 1.32 | 1.62 | 0.81 | 1 | 2 | 3 | 4 | 7 | 11 | 59 |
| | 4 | 78603 | 72331 | 0.92 | 1.25 | 0.79 | 1 | 1 | 2 | 3 | 5 | 8 | 35 |
| | 5 | 89923 | 91635 | 1.02 | 1.34 | 0.81 | 1 | 2 | 3 | 4 | 6 | 9 | 18 |
| | 6 | 92145 | 86016 | 0.93 | 1.24 | 0.80 | 1 | 1 | 3 | 3 | 5 | 8 | 20 |
| | 7 | 86701 | 80714 | 0.93 | 1.25 | 0.79 | 1 | 1 | 3 | 3 | 5 | 9 | 15 |
| | 8 | 91500 | 92147 | 1.01 | 1.33 | 0.79 | 1 | 2 | 3 | 4 | 6 | 10 | 22 |
| SC | 3 | 107293 | 171777 | 1.60 | 1.99 | 0.83 | 1 | 2 | 4 | 5 | 9 | 16 | 33 |
| | 4 | 78507 | 85004 | 1.08 | 1.57 | 0.84 | 1 | 2 | 3 | 4 | 7 | 13 | 41 |
| | 5 | 89823 | 95316 | 1.06 | 1.49 | 0.84 | 1 | 2 | 3 | 4 | 7 | 12 | 29 |
| | 6 | 92034 | 91799 | 1.00 | 1.39 | 0.82 | 1 | 1 | 3 | 4 | 6 | 11 | 23 |
| | 7 | 86556 | 78220 | 0.90 | 1.33 | 0.82 | 0 | 1 | 2 | 3 | 6 | 10 | 24 |
| | 8 | 91347 | 82516 | 0.90 | 1.30 | 0.81 | 0 | 1 | 2 | 3 | 6 | 10 | 27 |
| SS | 3 | 106857 | 218141 | 2.04 | 2.73 | 0.89 | 1 | 3 | 5 | 7 | 12 | 27 | 57 |
| | 4 | 78167 | 79208 | 1.01 | 1.61 | 0.86 | 1 | 1 | 3 | 4 | 7 | 14 | 49 |
| | 5 | 89503 | 104577 | 1.17 | 1.70 | 0.86 | 1 | 2 | 3 | 4 | 7 | 15 | 50 |
| | 6 | 91744 | 92117 | 1.00 | 1.62 | 0.85 | 1 | 1 | 3 | 4 | 7 | 15 | 54 |
| | 7 | 86183 | 83809 | 0.97 | 1.70 | 0.87 | 0 | 1 | 3 | 4 | 6 | 15 | 61 |
| | 8 | 90927 | 90581 | 1.00 | 1.64 | 0.86 | 1 | 1 | 3 | 4 | 6 | 17 | 50 |

5

**Table 3. Number of Schools Flagged for Erasure and WTR Erasure and WTR Analysis**

| Grade | Total Number of Schools | All Erasure Analyses | | Wrong-to-Right Erasure Analyses | |
|---|---|---|---|---|---|
| | | Number of Schools Flagged for at Least One Content Area | Number of Schools Flagged for All Content Areas | Number of Schools Flagged for at Least One Content Area | Number of Schools Flagged for All Content Areas |
| 3 | 1112 | 117 | 0 | 78 | 0 |
| 4 | 912 | 84 | 1 | 71 | 1 |
| 5 | 1034 | 99 | 0 | 66 | 0 |
| 6 | 516 | 78 | 2 | 53 | 0 |
| 7 | 492 | 67 | 2 | 54 | 0 |
| 8 | 492 | 67 | 4 | 57 | 1 |

6

Table 4. The number/percentage of schools that had zero flags for all erasures and wrong-to-right erasures

| | English/Language Arts | | | Mathematics | | | Science | | | Social Studies | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grade | No. of Schools | No. Of Schools with Zero Flags | % of Schools with Zero Flags | No. of Schools | No. Of Schools with Zero Flags | % of Schools with Zero Flags | No. of Schools | No. Of Schools with Zero Flags | % of Schools with Zero Flags | No. of Schools | No. Of Schools with Zero Flags | % of Schools with Zero Flags |
| 3 | 1112 | 1046 | 94% | 1111 | 1069 | 96% | 1111 | 1071 | 96% | 1111 | 1056 | 95% |
| 4 | 910 | 876 | 96% | 910 | 872 | 96% | 910 | 873 | 96% | 909 | 866 | 95% |
| 5 | 1018 | 973 | 96% | 1033 | 998 | 97% | 1030 | 995 | 97% | 1031 | 982 | 95% |
| 6 | 514 | 480 | 93% | 514 | 479 | 93% | 511 | 484 | 95% | 512 | 468 | 91% |
| 7 | 490 | 466 | 95% | 489 | 465 | 95% | 489 | 462 | 94% | 489 | 448 | 92% |
| 8 | 489 | 459 | 94% | 491 | 469 | 96% | 488 | 457 | 94% | 490 | 449 | 92% |

Table 5. The number/percentage of schools that had less than 1% of the classes flagged for all erasures and wrong-to-right erasures

| English/Language Arts | | | Mathematics | | | Science | | | Social Studies | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| No. of Schools | No. of Schools with <1% Flag Across Grades | % of Schools with <1% Flag Across Grades | No. of Schools | No. of Schools with <1% Flag Across Grades | % of Schools with <1% Flag Across Grades | No. of Schools | No. of Schools with <1% Flag Across Grades | % of Schools with <1% Flag Across Grades | No. of Schools | No. of Schools with <1% Flag Across Grades | % of Schools with <1% Flag Across Grades |
| 1721 | 1674 | 97% | 1721 | 1695 | 98% | 1718 | 1678 | 98% | 1720 | 1650 | 96% |

7

## Appendix D:  2015 DRC EOC Executive Report

**Erasure Analysis**

**Submitted by DRC**

**December 2015**

With the high-stakes nature of large-scale assessments such as the Milestones End of Course (EOC), there are times when students' responses, and hence their scores, may not be a true representation of their own abilities. Various activities may take place, such as a student copying from another student's paper, students receiving inappropriate assistance before or during testing, or students' responses altered after testing. To maintain the integrity of the Milestones EOC and the validity of the results, it is important that any such instances be discovered.

The present study investigated student responses on the Ninth Grade Literature, American Literature, Analytic Geometry, Coordinate Algebra, Biology, Physical Science, Economics/Business/Free Enterprise, and United States History tests of the 2015 Spring Milestones EOC that a) were erased and b) changed from a wrong answer to a right answer (wrong-to-right).

It should be emphasized that results from the erasure analyses performed in 2015 should only be used to identify potential problems within individual classrooms. That is, these types of analyses must be supported by additional, collateral information before conclusions regarding any improprieties are reached.

### Scanning Operations

The GA Milestones EOC answer documents were processed using high speed 5000i optical scanners which reliably captured document images and optical mark read data. The sophisticated proprietary scoring software system, specifically Optical Mark Recognition (OMR) software, reviews the integrity of each batch of documents scanned according to pre-defined guidelines and services.

The OMR software provides a mechanism for identifying multiple-marks and identification of erasures for scanned data. The basis of the erasure analysis is to count erasures for multiple-choice items where two or more responses have been made with a specified intensity. Erasure analyses provide a mechanism to differentiate between three kinds of answer changes: a) wrong-to-wrong, b) right-to-wrong and c) wrong-to-right. Capturing the frequency of answer changes from wrong-to-right can be useful for identifying potential instances of cheating at the student level. Erasure analyses results can be grouped to tentatively identify problems at the classroom and school levels.

### Method

The basis for the erasure analysis is to count erasures in items where an answer choice was erased and replaced with another answer choice. Often the data captured is useful for

identifying cases of cheating. During erasure analysis, two sets of erasures were analyzed: all erasures and wrong-to-right erasures where an incorrect answer choice was erased and replaced with the correct answer choice. Please note that, for the erasure analyses, all items (both operational and field-test) were included, as all field-test items were embedded in the Milestones EOC.

The basic idea underlying the procedure is a statistical test of the null hypothesis ($H_0$) that the mean number of erasures for a class constitutes a random sample from the state distribution of erasures. The hypothesis is tested against the (right-sided) alternative ($H_1$) that the mean number is too high to be explained by random sampling. Classes for which $H_0$ has to be rejected are flagged for further scrutiny. A well-known central limit theorem in statistics tells us that the sampling distribution of the mean number of erasures for class $i$ ($m_i$) is asymptotically normal with mean and standard deviation (SD)

$$mean(m_i) = \mu \qquad\qquad (1)$$

$$SD(m_i) = \frac{\sigma}{\sqrt{n_i}} \qquad\qquad (2)$$

where $n_i$ and $m_i$ denote the size and mean number of erasures for class $i$, respectively. In addition, $\mu$ and $\sigma$ denote the mean and the SD of the distribution of the number of erasures of the population of individual students in the state of Georgia.

The classes were flagged if their $m_i$ was larger than $\mu + 5\frac{\sigma}{\sqrt{n_i}}$. Statistically, the flagging criterion set at or above $5\sigma$ is conservative. The standard normal table shows that under random sampling the (asymptotic) probability of a sample mean being more than five SDs above the population mean is around 0.0000003. However, rejection of $H_0$ only tells us that the observed mean number of erasures is unlikely to be the result of random sampling.

It is evident in the formula that the class flagging criterion for each class is adjusted for the number of test takers in a classroom. For example, if the state mean and SD of erasure count are 1.73 and 2.11, respectively, the flagging criterion for a class size of 20 is adjusted to 4.11 ($1.73 + 5\frac{2.11}{\sqrt{20}} = 4.11$ ).

This adjustment ensures that the flagging criterion is equally stringent for classes with considerably different numbers of test takers. In addition, minimizing the probability of false positive (Type I) errors in this statistical test is crucial in this analysis.

<div align="center">

**Results**

</div>

Table 1 reports the state summary of erasure counts. The table includes the number of students, the total number of all types of erasures, the mean and the SD of all types of erasures, the correlation between all erasures and wrong-to-right erasures, the number of erasures at the 50th, 75th, 90th, 95th, 99th, and 99.9th percentiles, and the maximum number of all types of erasures. The mean number of erasures across all courses ranged from 1.46 to 2.37 for the 2015

2

Spring Milestones EOC. In other words, approximately 1 to 2 answer changes were made per student answer sheet on average. The erasure count at specific percentile points (50th, 75th, 90th, 95th, 99th, and 99.9th) is also reported. The erasure count at the 95th percentile point was between 5 and 7.

Table 2 reports the state summary of wrong-to-right erasure counts. The table includes the number of students, the number of wrong-to-right erasures, the mean and the SD of wrong-to-right erasures, the correlation between all erasures and wrong-to-right erasures, the number of wrong-to-right erasure at the 50th, 75th, 90th, 95th, 99th, and 99.9th percentiles, and the maximum number of wrong-to-right erasures. As can be expected, the mean wrong-to-right erasure count and the count at the specific percentile points were lower than those obtained from all erasure counts. The mean number of wrong-to-right erasures ranged from 0.78 to 1.10 for the 2015 Spring Milestones EOC. In other words, approximately 1 wrong-to-right answer change was made per student answer sheet on average. The wrong-to-right erasure count at specific percentile points (50th, 75th, 90th, 95th, 99th, and 99.9th) is also reported. The wrong-to-right erasure count at the 95th percentile point was between 3 and 4.

Table 3 presents a summary of the number of schools flagged for total erasures and wrong-to-right erasures based on Milestones EOC tests. Table 4 presents a summary of all schools with at least one class taking the Milestones EOC for at least one subject.

### Discussion

With respect to the erasure analyses, the following caveats are always applicable:

1. The normal distribution holds only for large classes; for smaller classes the result is approximate.

2. Rejection of $H_0$ does not necessarily imply cheating. Alternative explanations are possible.

3. The flagging criterion should thus be taken as a stimulus to look for additional evidence and find out what happened in the school.

This erasure analysis is considered a check for unusual numbers of erasures to student responses. Without additional layers added to the analysis, this kind of check only addresses the possibility, not the certainty, of teachers or administrators altering the responses of students. The 2015 erasure analyses represent an important step in helping to maintain the integrity of future administrations of the Milestones EOC.

**Table 1. State Summary Statistics for All Types of Erasure (ERA) Counts by Course**

| Course | N | No. of Erasures | Mean | SD | Correlation between ERA and WTR | Number of Erasure by Percentiles | | | | | | Max |
|--------|-----|-----|------|------|------|----|----|----|----|----|------|-----|
| | | | | | | 50 | 75 | 90 | 95 | 99 | 99.9 | |
| AME | 32434 | 47467 | 1.46 | 1.81 | 0.82 | 1 | 2 | 4 | 5 | 8 | 15 | 40 |
| ANA | 28845 | 57222 | 1.98 | 2.42 | 0.81 | 1 | 3 | 5 | 6 | 11 | 20 | 33 |
| BIO | 36278 | 72576 | 2.00 | 2.53 | 0.82 | 1 | 3 | 5 | 7 | 12 | 21 | 46 |
| COO | 37290 | 79199 | 2.12 | 2.51 | 0.80 | 1 | 3 | 5 | 7 | 11 | 22 | 37 |
| ECO | 7751 | 17561 | 2.27 | 2.73 | 0.84 | 1 | 3 | 5 | 7 | 13 | 22 | 27 |
| NTH | 44679 | 65711 | 1.47 | 1.87 | 0.82 | 1 | 2 | 4 | 5 | 8 | 14 | 49 |
| PHY | 15110 | 35786 | 2.37 | 2.74 | 0.83 | 2 | 3 | 6 | 7 | 13 | 20 | 39 |
| USH | 29120 | 66208 | 2.27 | 2.62 | 0.82 | 2 | 3 | 5 | 7 | 12 | 21 | 34 |

**Table 2. State Summary Statistics for Wrong-to-Right (WTR) Erasure Counts by Course**

| Course | N | No. of Erasures | Mean | SD | Correlation between ERA and WTR | Number of Erasure by Percentiles | | | | | | Max |
|--------|-----|-----|------|------|------|----|----|----|----|----|------|-----|
| | | | | | | 50 | 75 | 90 | 95 | 99 | 99.9 | |
| AME | 32434 | 25256 | 0.78 | 1.14 | 0.82 | 0 | 1 | 2 | 3 | 5 | 9 | 15 |
| ANA | 28845 | 25570 | 0.89 | 1.33 | 0.81 | 0 | 1 | 2 | 3 | 6 | 11 | 22 |
| BIO | 36278 | 34825 | 0.96 | 1.41 | 0.82 | 1 | 1 | 3 | 4 | 6 | 11 | 23 |
| COO | 37290 | 36403 | 0.98 | 1.37 | 0.80 | 1 | 1 | 3 | 3 | 6 | 11 | 21 |
| ECO | 7751 | 8528 | 1.10 | 1.56 | 0.84 | 1 | 2 | 3 | 4 | 7 | 13 | 18 |
| NTH | 44679 | 35378 | 0.79 | 1.16 | 0.82 | 0 | 1 | 2 | 3 | 5 | 8 | 14 |
| PHY | 15110 | 16377 | 1.08 | 1.51 | 0.83 | 1 | 2 | 3 | 4 | 7 | 12 | 20 |
| USH | 29120 | 31795 | 1.09 | 1.51 | 0.82 | 1 | 2 | 3 | 4 | 7 | 12 | 22 |

4

**Table 3. Number of Schools Flagged for Erasure and WTR Erasure and WTR Analysis**

| Course | Total Number of Schools | All Erasure Analyses | | Wrong-to-Right Erasure Analyses | |
|---|---|---|---|---|---|
| | | Number of Schools Flagged | Percent of Schools Flagged | Number of Schools Flagged | Percent of Schools Flagged |
| AME | 269 | 4 | 1.49 | 1 | 0.37 |
| ANA | 254 | 6 | 2.36 | 6 | 2.36 |
| BIO | 263 | 6 | 2.28 | 2 | 0.76 |
| COO | 317 | 7 | 2.21 | 6 | 1.89 |
| ECO | 175 | 3 | 1.71 | 0 | 0.00 |
| NTH | 302 | 7 | 2.32 | 3 | 0.99 |
| PHY | 240 | 3 | 1.25 | 4 | 1.67 |
| USH | 265 | 5 | 1.89 | 5 | 1.89 |

**Table 4. Number of Schools Flagged (WTR) in any Course for Milestones EOC**

| Course | Total Number of Schools | Number of Schools Flagged (WTR) | % of Schools Flagged (WTR) | Number of Schools Not Flagged (WTR) | % of Schools Not Flagged (WTR) |
|---|---|---|---|---|---|
| Paper Tests | 416 | 23 | 5.53 | 393 | 94.47 |

5