

**Statewide Analysis: Spring 2014 EOCT Courses***Presented to the State Board of Education on February, 2015***Rationale for this Investigation**

The Governor's Office of Student Achievement (GOSA) is charged with auditing and inspecting schools and Local Education Agencies (O.C.G.A. § 20-14-26). An analysis of the 2014 Spring EOCT answer documents conducted by the state's vendor, NCS Pearson, Inc., showed an unusually high number of answers changed from wrong to right (WTR) in some classrooms. Based on a conservative criterion for identifying unusual results, GOSA makes the recommendations in this report to help eliminate test misconduct and to help students adversely affected where applicable.

Because important decisions for individual students and for schools are based on EOCT data, it is vital that scores are an accurate representation of students' knowledge.

Purpose of the End-of-Course Test (EOCT)

The EOCT is a standardized assessment administered to students enrolled in high school core content courses. It is designed to measure how well students ascertain the knowledge and skills within the state's curriculum, the Georgia Performance Standards (GPS) outlined in the Common Core GPS and provide diagnostic information to help students identify strengths and areas of need in learning.

Executive Summary**Erasure Analysis**

The state's test vendor for the EOCT, NCS Pearson, Inc., conducted an erasure analysis on the Spring 2014 answer documents for the high school core content courses (Mathematics II, GPS Geometry, Analytic Geometry, Coordinate Algebra, United States History, Economics/Business/Free Enterprise, Biology, Physical Science, Ninth Grade Literature, and American Literature). The analysis was designed to identify any classroom in which answers were changed from wrong to right more frequently compared to the rest of the state test population in each course.

Using a professional grade scanner, Pearson scanned the answer sheets to determine the total number of erasures and the total number of wrong-to-right (WTR) changes on each document. Pearson then aggregated those results at the classroom level. Any classroom in which the number of WTR changes was 3 standard deviations (SD) or more (adjusted for class size) above the state average for that particular course was "flagged" as having an unusually high number of WTR changes.

Erasure Analysis Results

The mean range of WTR on the Spring 2014 analysis was 0.186 to 0.356 across subjects. These values indicate the vast majority, over 90% of students, did not have an answer changed from wrong to right. Approximately 99% of the schools either had less than 10% of their classrooms flagged; or fewer than 5 of their classrooms flagged. However, the analysis indicates there are still some classrooms that show an unusually high number of wrong answers changed to right answers.

Spring 2014 EOCT Erasure Analysis

Rationale for this Report

The Governor's Office of Student Achievement (GOSA) serves as the reporting and accountability agency for education in Georgia. As such, it is charged by law (O.C.G.A § 20-14-26) with auditing and inspecting schools and Local Education Agencies. As the vendor for the delivery of the End-of-Course Tests (EOCT), Pearson is providing services to GOSA to complete an Erasure Analysis for the Spring 2014 EOCT. This report is a summary of the comprehensive analysis completed on the Spring 2014 EOCT main administration.

EOCT Assessment Overview

The A+ Educational Reform Act of 2000, O.C.G.A. §20-2-281, mandates that the State Board of Education adopt end-of-course assessments in grades nine through twelve for core subjects to be determined by the State Board of Education. The EOCT is administered upon the completion of Analytic Geometry, Mathematics II, GPS Geometry, Coordinate Algebra, United States History, Economics, Biology, Physical Science, Ninth Grade Literature, and American Literature (a total of 10 EOCT). It is designed to measure both the effectiveness of classroom instruction at the school, system, and state levels and the strengths and areas of need in learning for examinees. A student's final grade in the applicable course is calculated as follows (State Board Rule 160-4-2-.13):

- For students enrolled in grade nine for the first time before July 1, 2011, the EOCT counts as 15% of their final grade
- For students enrolled in grade nine for the first time on, or after, July 1, 2011, the EOCT counts as 20% of their final grade

Georgia school systems have the option to administer the test in one of three ways:

- Paper and pencil administration with answer documents sent to Pearson for scoring
- Paper and pencil administration with answer documents scanned locally with software developed by Pearson
- Online administration with electronic responses scored at Pearson

The EOCT is part of Georgia's high school accountability assessment and is used as part of the College and Career Readiness Performance Index (CCRPI). Important decisions for individual students and for schools are based on EOCT data. Therefore, it is critical that reported scores are an accurate representation of students' knowledge.

Erasure Analysis

Pearson conducted an erasure analysis for the Spring 2014 EOCT main administration for assessments completed via paper and pencil. The analysis was conducted for EOCT in Analytic Geometry, Mathematics II, GPS Geometry, Coordinate Algebra, United States History, Economics, Biology, Physical Science, Ninth Grade Literature, and American Literature. The purpose was to identify classrooms where item responses were changed more frequently when compared to the typical EOCT classroom for the state test population.

Erasure Capture Method

Pearson uses optical mark scanners (OMR) to capture data from the scannable forms used on the EOCT. Scanners have the ability to discern between pre-printed coding and respondent markings using a 16-level mark discrimination system. An erasure, for paper testing, is determined by the following criteria: The highest intensity mark on the answer document is automatically classified as the examinee's response while the discernible mark with second darkest intensity is classified as the erasure. All scanned EOCT answer documents were analyzed using the mark discrimination system to determine responses that changed from wrong to right (WTR), responses that changed from right to wrong, responses that changed from wrong to wrong, and total erasures on each answer document.

Statistical Method

The method used to analyze the erasure data uses state and classroom population mean and variance adjusted for class size. The flagging procedure was applied to total and WTR erasures. The erasure analysis utilized data from all test items, including embedded field test items.

The statistical test used for flagging is based on a test of the null hypothesis (H_0) that the mean number of erasures for a class is drawn from a random sample from the state distribution of erasures for a class. The alternative hypothesis (H_1) is that the mean number of erasures for a class is too high to be the result of a random sample. Classes that are flagged due to the rejection of H_0 should be further analyzed to see if there is a non-random explanation for the flag.

The central limit theorem holds that the sampling distribution of a mean number of erasures for class c (m_c) is asymptotically normal with mean and standard deviation

$$\text{Mean}(m_c) = \mu \quad (1)$$

$$\text{SD}(m_c) = \frac{\sigma}{\sqrt{n_c}} \quad (2)$$

with n_c denoting the size of class c and m_c denoting the mean number of erasures for class c . Additionally, μ and σ denote the mean and standard deviation of the number of erasures of the population of examinees taking the EOCT in Georgia.

Classes were flagged if the m_c was larger than $\mu + 3 \frac{\sigma}{\sqrt{n_c}}$. Dividing the standard deviation by the square-root of n (Equation 2) allows the statistical test to be sensitive to different class sizes. For example, if the state mean and standard deviation are 2.34 and 3.65 respectively, the flagging criterion for a class size of 16 would be $\left(2.34 + 3 \frac{3.65}{\sqrt{16}}\right) = 5.08$, while the flagging criterion for a class size of 36 would be $\left(2.34 + 3 \frac{3.65}{\sqrt{36}}\right) = 4.17$.

The flagging criterion was set at 3σ to minimize the probability of false positive errors (Type I) in the statistical analysis. Under a random sampling of a normally distributed variable, the standard normal table shows that the probability of a sample mean being more than three standard deviations above the population mean is approximately 0.001. Rejection of H_0 only shows that the observed mean number of erasures for that particular class is unlikely to be the result of random sampling.

Erasure Analysis Results

Key initial findings from the erasure analysis include:

- The mean number of WTR erasures ranged from 0.186 to 0.356 across the 10 EOCT subjects. On average, approximately 0 to 1 wrong answers were erased and then correctly answered per

examinee per answer sheet.

- 69.18% of schools were not flagged in any subjects for WTR erasure.
- 50 of the 558 schools (8.96%) had at least one paper EOCT flagged for a higher than expected number of WTR erasures for at least 10% of their classrooms.
 - Of those 50 schools, 28 schools had only one classroom flagged, 14 schools had two classrooms flagged, and 15 schools had more than two classrooms flagged (ranging from four to thirty classrooms flagged).
- 28 of 558 schools (5.02%) had one classroom flagged for WTR erasures and at least 10% of their classrooms were flagged for WTR erasures.
- Approximately 99% of the schools had either less than 10% of their classrooms flagged or fewer than five of their classrooms flagged for a higher than expected number of WTR erasures.

Spring 2014 EOCT Main Administration Erasure Analysis Summary Tables

State Summary Statistics for Total Erasures by EOCT Subject

Table 1 displays a summary of the erasure data for EOCT tests administered. The table consists of subject area, total number of examinees (N), total number of erasures, the mean and standard deviation for total number of erasures, the correlation between the total number of erasures and WTR erasures, the number of erasures by percentile and the maximum number of erasures. The mean number of erasures ranged from 0.450 to 0.739 across the 10 EOCT subjects. Approximately 0 to 1 response was erased per examinee per answer sheet. At the 90th percentile, the erasure count was 1 to 2 for all subjects, which means that 90% of examinees had less than that number of erasures for the respective EOCT subject.

Table 1. State Summary Statistics for Total Erasures by EOCT Subject

Subject Area	N	Number of Erasures	Mean	Standard Deviation	Correlation with WTR Erasures	Number of Erasures by Percentiles						Maximum Number of Erasures
						50	75	90	95	99	99.9	
Analytic Geometry	59881	30738	0.513	1.229	0.804	0	1	2	3	6	11	36
Mathematics II	2228	1554	0.697	1.537	0.769	0	1	2	3	7	15	18
GPS Geometry	283	209	0.739	1.682	0.841	0	1	2	4	9	16	16
Coordinate Algebra	74489	33490	0.45	1.113	0.789	0	0	1	2	5	10	29
US History	43711	26370	0.603	1.479	0.866	0	1	2	3	7	14	39
Economics	21797	14723	0.675	1.655	0.863	0	1	2	3	7	18	38
Biology	56965	40031	0.703	1.55	0.828	0	1	2	3	7	15	39
Physical Science	34571	25106	0.726	1.548	0.846	0	1	2	4	7	15	26
9th Grade Lit.	63794	39029	0.612	1.34	0.848	0	1	2	3	6	12	45
American Lit.	48207	32134	0.667	1.408	0.863	0	1	2	3	7	12	32

State Summary Statistics for WTR Erasures by EOCT Subject

Table 2 displays a summary of the **WTR erasures** for EOCT tests administered. The table consists of subject area, total number of examinees (N), total number of WTR erasures, the mean and standard deviation for WTR erasures, the correlation between the total number of erasures and WTR erasures, the number of WTR erasures by percentile, and the maximum number of WTR erasures. The mean number of WTR erasures ranged from 0.186 to 0.356 across the 10 EOCT subjects. Approximately 0 to 1 wrong responses were erased and then correctly answered per examinee per answer sheet. At the 90th percentile, the WTR erasure count was 1 for all subjects, which means that 90% of examinees had less than that number of WTR erasures for the respective EOCT subject.

Table 2. State Summary Statistics for Total WTR Erasures by EOCT Subject

Subject Area	N	Number of WTR Erasures	Mean	Standard Deviation	Correlation with Erasures	Number of Erasures by Percentiles						Maximum Number of WTR Erasures
						50	75	90	95	99	99.9	
Analytic Geometry	59881	12869	0.215	0.632	0.804	0	0	1	1	3	6	24
Mathematics II	2228	553	0.248	0.661	0.769	0	0	1	1	3	5	11
GPS Geometry	283	75	0.265	0.686	0.841	0	0	1	1	4	5	5
Coordinate Algebra	74489	13855	0.186	0.573	0.789	0	0	1	1	3	5	14
US History	43711	13245	0.303	0.859	0.866	0	0	1	2	4	8	26
Economics	21797	7031	0.323	0.91	0.863	0	0	1	2	4	10	21
Biology	56965	17378	0.305	0.802	0.828	0	0	1	2	4	7	20
Physical Science	34571	12212	0.353	0.862	0.846	0	0	1	2	4	7	16
9th Grade Lit.	63794	20480	0.321	0.807	0.848	0	0	1	2	4	7	18
American Lit.	48207	17169	0.356	0.866	0.863	0	0	1	2	4	7	19

Number of Schools Flagged for Erasure and WTR Analysis

Table 3 displays a summary of the number of schools flagged for **total erasures and WTR erasures based** on EOCT tests. The table consists of subject area, total number of schools, number of schools flagged, and percent of schools flagged for both total erasures and WTR erasures.

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

Subject Area	Total Number of Schools	Erasures		WTR	
		Number of Schools Flagged	Percent of Schools Flagged	Number of Schools Flagged	Percent of Schools Flagged
Analytic Geometry	337	49	14.54	37	10.98
Mathematics II	172	9	5.23	6	3.49
GPS Geometry	24	0	0.00	1	4.17
Coordinate Algebra	485	77	15.88	70	14.43
US History	314	46	14.65	48	15.29
Economics	262	30	11.45	24	9.16
Biology	350	51	14.57	40	11.43
Physical Science	337	34	10.09	33	9.79
9th Grade Lit.	396	70	17.68	57	6.41
American Lit.	320	68	10.00	44	10.00

Number of Schools Flagged (WTR) in Any Subject Area for EOCT

Table 4 displays a summary of all schools with at least one class taking the EOCT for at least one subject. The table consists of the following columns: total number of schools, number of schools flagged and percent of schools flagged for WTR, number of schools not flagged, and percent of schools not flagged for WTR erasures. Table 4 shows that 68.18% of schools were not flagged in any subjects for EOCT tests.

Table 4. Number of Schools Flagged (WTR) in Any Subject Area for EOCT

	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	558	172	30.82	386	69.18

Discussion

With the high-stakes nature of large-scale assessments such as the EOCT, there are times when examinee's scores may not be a true representation of his or her own abilities. This may occur due to an examinee copying from another examinee's paper, an examinee receiving inappropriate assistance before or during testing from a variety of sources, or an examinee's responses being altered after testing. To maintain the validity of the EOCT results, it is important that occurrences, such as those previously mentioned, be discovered and identified.

It must be emphasized that the erasure analyses should only be considered as an initial step for checking a class with higher numbers of erasures than the state average. Flagging a class does not necessarily suggest improper activities. There are many potential sources of variances and alternative explanations are possible. Flagging should be an indicator to seek additional evidence to identify a possible problem within a class (and extended to a school or a district). Therefore, further investigation is imperative. The erasure analyses for Spring 2014 EOCT main administration is one component to uphold the integrity of the EOCT program and the assessment process.

These erasure analyses should only be used to identify *potential* problems within individual classrooms. These analyses must be confirmed by additional evidence before any conclusions regarding improprieties can be reached. In addition, when the class size is small, for example with 10 or fewer students, the erasure analysis results are only approximate and should be viewed with caution.