



Statewide Analysis: Spring 2013 CRCT Grades 3-8 in Reading, English-Language Arts, and Mathematics

Presented to the State Board of Education on February 20, 2014

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). A comprehensive analysis of the 2013 Spring CRCT answer documents conducted by the state's vendor, CTB McGraw-Hill, 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 CRCT data, it is vital that scores provide an accurate representation of students' knowledge.

Purpose of the Criterion Referenced Competency Test (CRCT)

The CRCT is a standardized assessment administered in 2013 to elementary and middle school students in Georgia. It is designed to measure how well students at each grade level have acquired the knowledge and skills within Georgia's performance/content standards outlined in the CCGPS/GPS (Common Core Georgia Performance Standards/Georgia Performance Standards).

Executive Summary

Erasure Analysis

The state's testing vendor for the CRCT, CTB-McGraw Hill, conducted an erasure analysis on 2013 answer sheets identical to those conducted in previous years since 2009. The analysis included every test-taker in grades 3-8 in Reading, Language Arts, and Math, and 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 grade and subject.

Using a professional grade scanner, CTB scanned the answer sheets to determine the total number of erasures and the total number of wrong-to-right (WTR) changes on each document. CTB 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 grade and subject was "flagged" as having an unusually high number of WTR changes. CTB then aggregated those results at the school level.

Erasure Analysis Results

The results of the 2013 analysis are markedly better than those of the 2009 and 2010 analyses and relatively comparable to those of the 2011 and 2012 analyses. In 2013, significantly fewer classrooms were flagged across the state, and those classroom flags were significantly smaller than flags seen in 2009 and 2010 analyses.

Although a continuous decline in the number of flagged classrooms across the state exists, the Spring 2013 analysis indicates that there are still some classrooms showing unusually high numbers of wrong-to-right answer changes in Reading, Language Arts and Math.

As in previous 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. In 2013, schools were categorized as follows:

- 93% of Georgia's elementary and middle schools fell into the "Clear" category (compared to 80% in 2009, 87% in 2010, 90% in 2011 and 94% in 2012),
- 5.4% fell into "Minimal concern" (compared to 10% in both 2009 and 2010, 7.4% in 2011 and 4.5% in 2012),
- 1.2% fell into "Moderate concern" (compared to 6% in 2009, 3% in 2010, 2.6% in 2011 and 1.4% in 2012), and
- 0.1% fell into "Severe concern" (compared to 4% in 2009, 0.5% in 2010, 0.2% in 2011 and 0.2% in 2012).

Erasure Analysis

Submitted by CTB-McGraw Hill

October 2013

With the high-stakes nature of large-scale assessments such as the CRCT, 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 CRCT and the validity of the results, it is important that any such instances be discovered.

The present study investigated student responses on the Reading, English/Language Arts, and Mathematics tests of the 2013 spring 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 2013 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 CRCT 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 specified intensity. 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 CRCT.

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 mean number of erasures for class i (m_i) is asymptotically normal with mean and standard deviation (SD)

$$\text{mean}(m_i) = \mu \quad (1)$$

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

where n_i and m_i denote the size and mean number of erasures for class i , respectively. In addition, μ and σ 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 $\mu + 3 \frac{\sigma}{\sqrt{n_i}}$. Statistically, the flagging

criterion set at or above 3σ is conservative. The standard normal table shows that under random sampling the (asymptotic) probability of a sample mean being more than three SDs above the population mean is around 0.001. 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 3.15 ($1.73 + 3 \frac{2.11}{\sqrt{20}} = 3.15$).

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.

Results

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 grades and subjects ranged from 1.27 to 2.78 for the 2013 spring CRCT. 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, 95th, 99th, and 99.9th) is also reported. The erasure count at the 95th percentile point was between 5 and 9.

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 0.74 to 1.84 for the 2013 spring CRCT. 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 99.9th) is also reported. The wrong-to-right erasure count at the 95th percentile point was between 3 and 6.

Separate reports were produced displaying the results of all erasure analyses and wrong-to-right erasure analyses for the 2013 spring. Tables 3 through 5 summarize all erasure analyses and wrong-to-right erasure analyses of the 2013 spring CRCT.

Table 3 presents the number of schools flagged across three content areas—Reading, English/Language Arts, and Mathematics—within each analysis of the spring CRCT. For each analysis, the number of schools was computed in two ways: flagged for at least one content area or flagged for all three content areas. The number/percentage of schools that had zero flags for all erasures and wrong-to-right erasures in Reading, English/Language Arts, and Mathematics 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 Reading, English/Language Arts, and Mathematics 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 2013 erasure analyses represent an important step in helping to maintain the integrity of future administrations of the CRCT.

Table 1. State summary statistics for all types of erasure (ERA) counts by content and grade

Content	Grade	N	No. of Erasures	Mean	SD	Correlation With WTR	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	3	124,477	219,641	1.76	2.18	0.84	1	2	4	6	10	18	40
	4	123,209	179,139	1.45	1.96	0.86	1	2	4	5	9	17	45
	5	123,497	200,152	1.62	2.03	0.85	1	2	4	5	9	16	43
	6	125,291	161,572	1.29	1.84	0.85	1	2	3	5	8	16	45
	7	125,341	158,607	1.27	1.80	0.85	1	2	3	5	8	15	43
	8	122,939	172,570	1.40	1.92	0.86	1	2	4	5	9	16	48
LA	3	124,700	275,585	2.21	2.60	0.87	1	3	5	7	12	20	51
	4	123,155	239,170	1.94	2.39	0.88	1	3	5	6	11	19	45
	5	123,309	222,328	1.80	2.30	0.88	1	3	5	6	10	18	58
	6	125,167	189,210	1.51	2.11	0.88	1	2	4	5	9	18	54
	7	125,025	180,571	1.44	2.04	0.88	1	2	4	5	9	17	46
	8	122,862	196,282	1.60	2.21	0.88	1	2	4	6	10	19	59
MA	3	125,095	347,474	2.78	2.95	0.88	2	4	6	8	13	22	48
	4	122,873	304,133	2.48	2.74	0.89	2	4	6	8	12	20	48
	5	122,473	290,167	2.37	2.69	0.88	2	3	6	7	12	21	43
	6	124,509	262,828	2.11	2.51	0.85	1	3	5	7	11	19	61
	7	124,179	253,735	2.04	2.51	0.87	1	3	5	7	12	20	50
	8	122,063	297,823	2.44	2.79	0.86	2	3	6	8	13	22	57

Table 2. State summary statistics for wrong-to-right (WTR) erasure counts by content and grade

Content	Grade	N	No. of Erasures	Mean	SD	Correlation With ERA	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	3	124,477	134,028	1.08	1.45	0.84	1	2	3	4	6	12	30
	4	123,209	110,767	0.90	1.34	0.86	0	1	2	3	6	11	41
	5	123,497	117,816	0.95	1.34	0.85	1	1	3	3	6	10	26
	6	125,291	95,387	0.76	1.21	0.85	0	1	2	3	5	10	32
	7	125,341	92,273	0.74	1.18	0.85	0	1	2	3	5	9	39
	8	122,939	103,361	0.84	1.29	0.86	0	1	2	3	6	10	27
LA	3	124,700	174,581	1.40	1.83	0.87	1	2	4	5	8	14	33
	4	123,155	154,169	1.25	1.73	0.88	1	2	3	4	8	14	39
	5	123,309	141,280	1.15	1.62	0.88	1	2	3	4	7	13	40
	6	125,167	114,883	0.92	1.42	0.88	0	1	3	4	6	12	32
	7	125,025	112,955	0.90	1.41	0.88	0	1	3	3	6	12	32
	8	122,862	118,316	0.96	1.50	0.88	0	1	3	4	7	13	37
MA	3	125,095	230,045	1.84	2.17	0.88	1	3	5	6	10	16	37
	4	122,873	203,959	1.66	2.04	0.89	1	2	4	6	9	15	35
	5	122,473	188,695	1.54	1.96	0.88	1	2	4	5	9	15	35
	6	124,509	150,645	1.21	1.62	0.85	1	2	3	4	7	13	30
	7	124,179	149,659	1.21	1.67	0.87	1	2	3	4	8	13	34
	8	122,063	175,347	1.44	1.83	0.86	1	2	4	5	8	15	37

Table 3. The number of schools flagged across three content areas (Spring CRCT)

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	1,258	283	18	240	14
4	1,255	291	23	229	8
5	1,256	272	26	229	15
6	606	220	39	141	13
7	594	215	29	145	6
8	587	205	37	151	14

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

Grade	Reading			English/Language Arts			Mathematics			Reading, English/Language Arts, and Mathematics		
	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	1,258	1,062	84%	1,258	1,090	87%	1,258	1,110	88%	1,258	884	70%
4	1,255	1,059	84%	1,254	1,093	87%	1,253	1,097	88%	1,255	884	70%
5	1,256	1,078	86%	1,254	1,074	86%	1,254	1,111	89%	1,256	916	73%
6	605	457	76%	603	464	77%	604	483	80%	606	354	58%
7	594	460	77%	594	476	80%	594	471	79%	594	350	59%
8	587	453	77%	587	460	78%	587	469	80%	587	361	61%

Table 5. The number and percentage of schools that had less than 1% of classes flagged for all erasures and wrong-to-right erasures across grades (Spring CRCT)

Reading			English/Language Arts			Mathematics			Reading, English/Language Arts, and Mathematics		
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
1,825	1,521	83%	1,825	1,536	84%	1,825	1,584	87%	1,825	1,806	99%