



## Statewide Analysis: Spring 2012 CRCT Grades 3-8 in Reading, English-Language Arts, and Mathematics

*Presented to the State Board of Education on February 20, 2013*

### 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 2012 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, OSA 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 2012 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 the state's curriculum, the Georgia Performance Standards.

### Executive Summary

#### Erasure Analysis

The state's testing vendor for the CRCT, CTB-McGraw Hill, conducted an erasure analysis on 2012 answer sheets identical to those conducted in 2009, 2010, and 2011. 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 2012 analysis are markedly better than those of the 2009 and 2010 analyses and relatively comparable to those of the 2011 analysis. In 2012, significantly fewer classrooms were flagged across the state, and those classroom flags were significantly smaller than flags seen in previous years.

However, the Spring 2012 analysis indicates that there are still some classrooms showing unusually high numbers of wrong-to-right answer changes in Reading, Language Arts and Math. At the school level, the percentage of classrooms flagged using the conservative criterion of 3 SD above the state average ranged from 0% to 33.3%, with a statewide average of 1.8%.

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 2012, schools were categorized as follows:

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

## **Erasure Analysis**

**Submitted by CTB-McGraw Hill**

**October 2012**

With the high-stakes nature of large-scale assessments such as the CRCT, there are times when student's 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 2012 spring that a) were erased and b) changed from wrong-to-right answers.

It should be emphasized that results from the erasure analysis performed in 2012 should only be used to identify potential problems within individual classrooms. That is, this type of analysis 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)

$$Mean(m_i) = \mu \quad (1)$$

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

where  $n_i$  and  $m_i$  denote the size and mean number of erasures for class  $i$ , respectively. In addition,  $\mu$  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.

It is evident in the formula for the state SD 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.

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  $3\sigma$  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.

## 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 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>, 95<sup>th</sup>, 99<sup>th</sup>, and 99.9<sup>th</sup> percentiles, and the maximum number of all types of erasures. The mean number of erasures across grades and subjects ranged from 1.32 to 2.80 for the 2012 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 (50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>, 95<sup>th</sup>, 99<sup>th</sup>, and 99.9<sup>th</sup>) is also reported. The erasure count at the 95<sup>th</sup> 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 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>, 95<sup>th</sup>, 99<sup>th</sup>, and 99.9<sup>th</sup> percentiles, and the maximum

number of wrong-to-right erasure. 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 0.80 to 1.87 for the 2012 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 the 95<sup>th</sup> 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 2012 spring. Tables 3 through 5 summarize all erasure analyses and wrong-to-right erasure analyses of the 2012 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 2012 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	Corr. With WTR	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	3	123,670	221,876	1.79	2.17	0.85	1	3	4	6	10	18	41
	4	124,090	180,725	1.46	1.95	0.86	1	2	4	5	9	16	44
	5	125,782	199,424	1.59	2.03	0.84	1	2	4	5	9	17	35
	6	125,149	165,324	1.32	1.80	0.85	1	2	3	5	8	15	40
	7	122,877	174,901	1.42	1.91	0.84	1	2	4	5	9	16	36
	8	121,488	187,135	1.54	2.02	0.87	1	2	4	5	9	17	47
LA	3	123,752	269,787	2.18	2.57	0.88	1	3	5	7	12	20	45
	4	123,611	242,766	1.96	2.41	0.88	1	3	5	6	11	19	50
	5	125,628	229,333	1.83	2.29	0.88	1	3	5	6	10	18	45
	6	125,055	182,716	1.46	2.05	0.88	1	2	4	5	9	18	58
	7	122,628	175,438	1.43	2.05	0.88	1	2	4	5	9	18	48
	8	121,601	216,400	1.78	2.37	0.89	1	2	4	6	11	21	60
MA	3	124,213	335,215	2.70	2.89	0.88	2	4	6	8	13	22	54
	4	123,435	346,177	2.80	2.97	0.89	2	4	6	8	13	22	68
	5	124,953	335,462	2.68	2.87	0.88	2	4	6	8	13	22	46
	6	124,303	280,011	2.25	2.63	0.84	2	3	5	7	12	21	54
	7	121,666	265,995	2.19	2.67	0.88	1	3	5	7	12	21	60
	8	120,887	338,728	2.80	3.06	0.86	2	4	6	9	14	25	51

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

Content	Grade	N	No. of Erasures	Mean	SD	Corr. With ERA	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	3	123,670	138,698	1.12	1.47	0.85	1	2	3	4	6	12	29
	4	124,090	113,422	0.91	1.35	0.86	0	1	2	3	6	11	34
	5	125,782	114,636	0.91	1.32	0.84	0	1	2	3	6	10	26
	6	125,149	99,637	0.80	1.21	0.85	0	1	2	3	5	9	26
	7	122,877	101,865	0.83	1.24	0.84	0	1	2	3	5	10	32
	8	121,488	112,638	0.93	1.39	0.87	0	1	3	3	6	12	28
LA	3	123,752	174,376	1.41	1.86	0.88	1	2	4	5	8	15	37
	4	123,611	160,616	1.30	1.77	0.88	1	2	3	5	8	14	32
	5	125,628	144,821	1.15	1.62	0.88	1	2	3	4	7	13	39
	6	125,055	114,875	0.92	1.44	0.88	0	1	3	4	6	12	47
	7	122,628	114,649	0.93	1.46	0.88	0	1	3	4	6	12	44
	8	121,601	137,444	1.13	1.67	0.89	1	2	3	4	7	14	47
MA	3	124,213	223,259	1.80	2.16	0.88	1	3	4	6	10	16	40
	4	123,435	230,478	1.87	2.21	0.89	1	3	5	6	10	17	52
	5	124,953	210,315	1.68	2.04	0.88	1	2	4	6	9	16	30
	6	124,303	153,841	1.24	1.65	0.84	1	2	3	4	7	13	28
	7	121,666	162,341	1.33	1.84	0.88	1	2	3	5	8	15	58
	8	120,887	188,050	1.56	1.93	0.86	1	2	4	5	9	15	33

**Table 3. The number of schools flagged across three content areas**

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	1,266	304	26	261	17
4	1,261	311	21	260	7
5	1,261	283	17	240	14
6	613	230	36	155	7
7	587	199	24	160	9
8	580	212	44	162	12

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

Grade	Reading			English/Language Arts			Mathematics			Reading, English/Language Arts, and Mathematics		
	No. of Schools	No. of Schools with zero flag	% of Schools with zero flag	No. of Schools	No. of Schools with zero flag	% of Schools with zero flag	No. of Schools	No. of Schools with zero flag	% of Schools with zero flag	No. of Schools	No. of Schools with zero flag	% of Schools with zero flag
3	1,266	1,062	84%	1,266	1,081	85%	1,266	1,100	87%	1,266	887	70%
4	1,260	1,045	83%	1,260	1,079	86%	1,261	1,112	88%	1,261	871	69%
5	1,259	1,075	85%	1,260	1,079	86%	1,261	1,122	89%	1,261	904	72%
6	612	478	78%	611	464	76%	608	478	79%	613	357	58%
7	587	460	78%	587	458	78%	587	475	81%	587	349	59%
8	579	426	74%	578	451	78%	579	444	77%	580	332	57%

**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**

Reading			English/Language Arts			Mathematics			Reading, English/Language Arts, and Mathematics		
No. of Schools	No. of Schools with <1% flag across grades	No. of Schools with <1% flag across grades	No. of Schools	No. of Schools with <1% flag across grades	No. of Schools with <1% flag across grades	No. of Schools	No. of Schools with <1% flag across grades	No. of Schools with <1% flag across grades	No. of Schools	No. of Schools with <1% flag across grades	No. of Schools with <1% flag across grades
1,833	1,500	82%	1,833	1,503	82%	1,833	1,596	87%	1,833	1,819	99%