



Statewide Analysis: Spring 2014 CRCT Grades 3-8 in Reading, English-Language Arts, and Mathematics

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). A comprehensive analysis of the 2014 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 2014 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 2014 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 three 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 2014 analysis provide evidence of consistency as the percentage of flagged classrooms had flattened over the last three years. Still, in 2014, significantly fewer classrooms were flagged across the state than in 2009.

The Spring 2014 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 2014, 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, 90% in 2011, 94% in 2012, and 93% in 2013),
- 4.1% fell into "Minimal concern" (compared to 10% in both 2009 and 2010, 7.4% in 2011, 4.5% in 2012, and 5.4% in 2013),
- 1.5% fell into "Moderate concern" (compared to 6% in 2009, 3% in 2010, 2.6% in 2011, 1.4% in 2012, and 1.2% in 2013), and
- 0.1% fell into "Severe concern" (compared to 4% in 2009, 0.5% in 2010, 0.2% in 2011, 0.2% in 2012, and 0.1% in 2013).

Erasure Analysis

Submitted by CTB-McGraw Hill

October 2014

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 2014 spring and retest CRCT 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 2014 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 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 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 the 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

Error! Objects cannot be created from

. Statistically, the flagging criterion set at or above 3σ is conservative. The

editing field codes.

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

(**Error! Objects cannot be created from editing field codes.**).

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 1.29 to 2.83 for the 2014 spring CRCT and from 1.67 to 2.74 for the 2014 retest 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 0.70 to 1.87 for the 2014 spring CRCT and from 0.63 to 1.09 for the 2014 retest 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 2014 spring and retest CRCT. Tables 3 through 5 summarize all erasure analyses and wrong-to-right erasure analyses of the 2014 spring CRCT. Tables 6 through 8 summarize all erasure analyses and wrong-to-right erasure analyses of the 2014 retest 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.

Table 6 presents the number of schools flagged across two content areas—Reading and Mathematics—within each analysis of retest CRCT. For each analysis, the number of schools was computed in two ways: flagged for at least one content area or flagged for both content areas. The number/percentage of schools that had zero flags for all erasures and wrong-to-right erasures in Reading and Mathematics is provided in Table 7. The number/percentage of schools that had less than 1% of classes flagged for all erasures and wrong-to-right erasures in Reading and Mathematics across grades is provided in Table 8.

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 2014 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 between ERA and WTR	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	3	126,351	237,508	1.88	2.25	0.82	1	3	4	6	10	18	41
	4	124,575	198,994	1.60	2.02	0.84	1	2	4	5	9	16	44
	5	123,279	191,995	1.56	2.02	0.83	1	2	4	5	9	16	49
	6	124,622	160,698	1.29	1.83	0.84	1	2	3	5	8	16	44
	7	127,125	164,376	1.29	1.85	0.83	1	2	3	5	8	16	44
	8	125,893	189,393	1.50	2.09	0.86	1	2	4	5	9	19	43
LA	3	126,609	272,556	2.15	2.59	0.85	1	3	5	7	12	21	54
	4	124,414	231,971	1.86	2.36	0.86	1	3	5	6	11	19	57
	5	123,181	203,032	1.65	2.19	0.87	1	2	4	6	10	19	60
	6	124,474	189,507	1.52	2.14	0.87	1	2	4	5	10	19	39
	7	126,851	165,861	1.31	1.97	0.86	1	2	3	5	9	17	60
	8	125,678	216,085	1.72	2.40	0.88	1	2	4	6	11	22	57
MA	3	127,076	359,501	2.83	3.04	0.88	2	4	7	9	14	23	58
	4	124,390	317,920	2.56	2.83	0.88	2	4	6	8	13	22	53
	5	122,602	297,140	2.42	2.70	0.87	2	3	6	8	12	21	70
	6	123,900	260,332	2.10	2.49	0.85	1	3	5	7	11	19	49
	7	126,236	264,708	2.10	2.54	0.84	1	3	5	7	12	21	68
	8	124,894	316,026	2.53	2.92	0.85	2	4	6	8	13	24	67
RD Retest	3	9,342	18,774	2.01	2.88	0.79	1	3	5	7	14	27	48
	5	6,599	11,051	1.67	2.31	0.79	1	2	4	6	10	21	28
	8	4,568	8,230	1.80	3.02	0.85	1	2	5	7	14	30	37
MA Retest	5	14,593	33,446	2.29	2.82	0.82	2	3	5	7	13	25	62
	8	21,798	59,758	2.74	3.18	0.82	2	4	6	8	15	28	48

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 between ERA and WTR	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	3	126,351	137,672	1.09	1.42	0.82	1	2	3	4	6	11	23
	4	124,575	120,571	0.97	1.35	0.84	1	1	3	3	6	10	31
	5	123,279	105,313	0.85	1.25	0.83	0	1	2	3	5	9	36
	6	124,622	89,016	0.71	1.15	0.84	0	1	2	3	5	9	29
	7	127,125	88,651	0.70	1.13	0.83	0	1	2	3	5	9	39
	8	125,893	115,769	0.92	1.40	0.86	0	1	3	3	6	12	24
LA	3	126,609	161,113	1.27	1.71	0.85	1	2	3	4	8	13	35
	4	124,414	141,592	1.14	1.59	0.86	1	2	3	4	7	12	38
	5	123,181	121,951	0.99	1.47	0.87	1	1	3	4	6	12	46
	6	124,474	111,162	0.89	1.39	0.87	0	1	2	3	6	11	29
	7	126,851	98,333	0.78	1.30	0.86	0	1	2	3	6	11	47
	8	125,678	128,523	1.02	1.61	0.88	1	1	3	4	7	14	37
MA	3	127,076	237,159	1.87	2.28	0.88	1	3	5	6	10	18	51
	4	124,390	202,465	1.63	1.99	0.88	1	2	4	5	9	15	35
	5	122,602	179,540	1.46	1.87	0.87	1	2	4	5	8	14	41
	6	123,900	144,857	1.17	1.58	0.85	1	2	3	4	7	12	33
	7	126,236	148,250	1.17	1.60	0.84	1	2	3	4	7	12	34
	8	124,894	178,537	1.43	1.84	0.85	1	2	4	5	8	14	36
RD Retest	3	9,342	6,792	0.73	1.23	0.79	0	1	2	3	5	11	18
	5	6,599	4,160	0.63	1.07	0.79	0	1	2	3	5	8	14
	8	4,568	3,246	0.71	1.34	0.85	0	1	2	3	6	10	19
MA Retest	5	14,593	14,488	0.99	1.43	0.82	1	1	3	4	6	11	23
	8	21,798	23,862	1.09	1.50	0.82	1	2	3	4	7	11	24

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	1250	328	19	274	8
4	1246	279	31	211	13
5	1249	315	22	230	9
6	597	217	29	164	9
7	588	191	23	139	5
8	586	218	41	174	21

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	1250	1024	82%	1250	1044	84%	1250	1080	86%	1250	829	66%
4	1245	1068	86%	1245	1077	87%	1246	1107	89%	1246	905	73%
5	1249	1053	84%	1247	1073	86%	1246	1089	87%	1249	862	69%
6	597	456	76%	597	464	78%	596	477	80%	597	345	58%
7	586	456	78%	585	479	82%	587	472	80%	588	356	61%
8	585	432	74%	586	446	76%	586	443	76%	586	337	58%

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
1810	1486	82%	1810	1517	84%	1810	1571	87%	1811	1791	99%

Table 6. The number of schools flagged across two content areas (Retest 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	1138	22	N/A	26	N/A
5	1209	46	3	47	1
8	574	30	3	29	3

Table 7. The number and percentage of schools that had zero flags for all erasures and wrong-to-right erasures (Retest CRCT)

Grade	Reading			Mathematics			Reading 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
3	1138	1105	97%	N/A	N/A	N/A	N/A	N/A	N/A
5	1099	1070	97%	1193	1149	96%	1209	1139	94%
8	508	495	97%	573	544	95%	574	536	93%

Table 8. The number and percentage of schools that had less than 1% of classes flagged for all erasures and wrong-to-right erasures across grades (Retest CRCT)

Reading			Mathematics			Reading and Mathematics		
No. of Schools	No. of Schools with <1% flags across grades	% of Schools with <1% flags across grades	No. of Schools	No. of Schools with <1% flags across grades	% of Schools with <1% flags across grades	No. of Schools	No. of Schools with <1% flags across grades	% of Schools with <1% flags across grades
1695	1665	98%	1704	1672	98%	1719	1716	100%