



Statewide Analysis: Spring 2011 CRCT Grades 3-8 in Reading, English-Language Arts, and Mathematics
Presented to the State Board of Education on April 5, 2012

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 2011 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 are an accurate representation of students' knowledge.

Purpose of the Criterion Referenced Competency Test (CRCT)

The CRCT is a standardized assessment administered in 2011 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 test vendor for the CRCT, CTB-McGraw Hill, conducted an erasure analysis on 2011 answer sheets identical to that done on 2009 and 2010 answer sheets. The analysis was conducted for 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 2011 analysis showed marked improvement from the analyses done in previous years on the 2009 and 2010 answer sheets. Significantly fewer classrooms were flagged across the state, and those flags were generally far smaller than flags seen previously.

However, the analysis indicates that there are still some classrooms that show an unusually high number of wrong answers changed to right answers on the grades 3-8 Spring 2011 CRCT 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 37.8%, with a statewide average of 1.1%.

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

2011 Erasure Analysis

Submitted by CTB-McGraw Hill

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 2011 spring CRCT that a) were erased and b) changed from wrong-to-right answers.

It should be emphasized that results from the erasure analyses performed in 2011 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. 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 (either the operational or 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

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

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

where n_i and m_i denote the size and mean number of erasures for class i , respectively. In addition, μ and σ denote the mean and the standard deviation 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 standard deviation 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 3σ is conservative. The standard normal table shows that under random sampling the (asymptotic) probability of a sample mean being more than three standard deviations 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 standard deviation (Std) 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 1.38 to 2.74. 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 standard deviation (Std) 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 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 0.81 to 1.82. 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 95th percentile point was between 3 and 6.

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 2011 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	Std	Corr. With WTR	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	3	125,505	215,902	1.72	2.14	0.84	1	2	4	6	10	18	44
	4	127,714	207,285	1.62	2.09	0.85	1	2	4	5	9	18	47
	5	126,213	208,510	1.65	2.08	0.84	1	2	4	5	9	17	48
	6	123,818	175,475	1.42	1.89	0.85	1	2	4	5	8	16	33
	7	122,027	174,647	1.43	1.90	0.84	1	2	4	5	8	16	42
	8	119,023	186,013	1.56	2.04	0.86	1	2	4	5	9	18	43
LA	3	125,201	263,262	2.10	2.48	0.88	1	3	5	7	11	19	57
	4	127,107	260,123	2.05	2.48	0.88	1	3	5	7	11	20	60
	5	126,088	233,916	1.86	2.29	0.88	1	3	5	6	10	18	40
	6	123,658	195,658	1.58	2.14	0.87	1	2	4	5	10	18	49
	7	121,996	168,265	1.38	1.96	0.88	1	2	4	5	9	17	43
	8	119,016	209,510	1.76	2.32	0.88	1	2	4	6	11	20	53
MA	3	125,641	304,106	2.42	2.72	0.89	2	3	6	8	12	21	56
	4	126,831	348,072	2.74	2.92	0.89	2	4	6	8	13	22	55
	5	125,139	336,872	2.69	2.87	0.88	2	4	6	8	13	21	43
	6	123,170	294,747	2.39	2.69	0.84	2	3	6	7	12	21	57
	7	121,255	237,734	1.96	2.43	0.87	1	3	5	6	11	20	42
	8	118,651	325,566	2.74	3.00	0.86	2	4	6	8	14	23	70

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

Content	Grade	N	No. of Erasures	Mean	Std	Corr. With ERA	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	3	125,505	133,575	1.06	1.43	0.84	1	2	3	4	6	12	25
	4	127,714	130,306	1.02	1.43	0.85	1	2	3	4	6	12	27
	5	126,213	115,762	0.92	1.32	0.84	0	1	2	3	6	11	30
	6	123,818	103,115	0.83	1.26	0.85	0	1	2	3	5	10	24
	7	122,027	98,482	0.81	1.22	0.84	0	1	2	3	5	9	29
	8	119,023	112,819	0.95	1.37	0.86	1	1	3	3	6	11	28
LA	3	125,201	173,778	1.39	1.80	0.88	1	2	4	5	8	14	32
	4	127,107	173,891	1.37	1.82	0.88	1	2	4	5	8	14	49
	5	126,088	148,728	1.18	1.63	0.88	1	2	3	4	7	13	29
	6	123,658	120,394	0.97	1.47	0.87	1	1	3	4	6	12	41
	7	121,996	104,472	0.86	1.37	0.88	0	1	2	3	6	12	26
	8	119,016	129,590	1.09	1.61	0.88	1	2	3	4	7	13	44
MA	3	125,641	201,583	1.60	2.02	0.89	1	2	4	5	9	16	45
	4	126,831	231,392	1.82	2.14	0.89	1	3	4	6	10	16	38
	5	125,139	206,814	1.65	2.01	0.88	1	2	4	5	9	15	37
	6	123,170	164,765	1.34	1.71	0.84	1	2	3	5	8	13	30
	7	121,255	145,680	1.20	1.66	0.87	1	2	3	4	7	13	28
	8	118,651	183,439	1.55	1.92	0.86	1	2	4	5	9	15	48

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	1272	317	23	274	13
4	1258	314	28	257	11
5	1262	319	25	262	11
6	592	215	32	159	5
7	572	209	37	158	18
8	575	220	48	176	19

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	1272	1053	83%	1272	1075	85%	1271	1106	87%	1272	868	68%
4	1257	1050	84%	1257	1082	86%	1258	1100	87%	1258	867	69%
5	1262	1062	84%	1262	1071	85%	1262	1084	86%	1262	874	69%
6	592	458	77%	590	460	78%	590	465	79%	592	347	59%
7	572	440	77%	571	438	77%	571	447	78%	572	334	58%
8	574	413	72%	574	435	76%	575	436	76%	575	323	56%

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
1834	1474	80%	1833	1483	81%	1834	1568	85%	1834	1816	99%