



Statewide Analysis: Spring 2010 CRCT Grades 1-8 in Reading, English-Language Arts, and Mathematics

Presented to the State Board of Education on February 9, 2011

Rationale for this Investigation

The Office of Student Achievement (OSA) is charged with auditing and inspecting schools and Local Education Agencies (O.C.G.A. § 20-12-24). A comprehensive analysis of the 2010 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 are an accurate representation of students' knowledge.

Purpose of the Criterion Referenced Competency Test (CRCT)

The CRCT is a standardized assessment administered in 2010 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.

CRCT results are used to determine whether schools have made Adequate Yearly Progress (AYP) as required by the *No Child Left Behind (NCLB) Act*.

Executive Summary

Erasure Analysis

The state's test vendor for the CRCT, CTB-McGraw Hill, conducted an erasure analysis on 2010 answer sheets identical to that done on 2009 answer sheets. The analysis was conducted for grades 1-8 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 2010 analysis showed marked improvement from that done a year ago on 2009 answer sheets. Significantly fewer classrooms were flagged across the state, and those flags were generally far smaller than flags seen last year.

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 1-8 Spring 2010 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 44%, with a statewide average of 2.4%.

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.

87% of Georgia's elementary and middle schools fell into the "Clear" category (compared to 80% in 2009); 10% fell into "minimal concern" (compared to 10% in 2009); 3% fell into "Moderate concern" (compared to 6% in 2009); and 0.5% fell into "Severe concern" (compared to 4% in 2009).

Erasure Analysis

Submitted by CTB-McGraw Hill

January 5, 2011

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 Spring 2010 CRCT that a) were erased and b) changed from wrong to right answers

It should be emphasized that results from erasure analyses performed in 2010 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 capture document images and optical mark read data. The sophisticated proprietary scoring software system WinScore reviews the integrity of each batch of documents scanned according to pre-defined guidelines and services. This flexible system reduces scanning/scoring time and provides a high degree of quality control.

The WinScore system 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 field test items were all embedded in 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 for erasure counts. The mean number of erasures across grades and subjects ranged from 1.33 to 3.08 for the 2010 spring CRCT. 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 for wrong-to-right erasure counts. 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 0.76 to 1.83 for the 2010 spring CRCT. The wrong-to-right erasure count at the 95th percentile point was between 3 and 6.

Table 3 summarizes all erasure analyses and wrong-to-right erasure analyses. Table 3 presents the number of schools flagged across three content areas –Reading, English/Language Arts, and Mathematics - within each analysis of 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.

Separate reports were produced displaying the results of all erasure analyses and wrong-to-right erasure analyses. The number/percentage of schools that had zero flags for all erasures and wrong-to-right erasures in Reading, English/Language Arts, and Mathematics and is provided in Table 4.

The number/percentage of schools that had less than 1% of 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 2010 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 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	1	127,630	235,173	1.84	2.09	0.86	1	3	5	6	9	14	41
	2	127,324	172,309	1.35	1.73	0.87	1	2	4	5	8	12	40
	3	130,076	239,835	1.84	2.23	0.83	1	3	4	6	10	19	45
	4	128,105	201,423	1.57	2.08	0.85	1	2	4	5	9	18	48
	5	126,221	220,367	1.75	2.16	0.84	1	2	4	6	10	18	46
	6	124,216	164,924	1.33	1.82	0.84	1	2	3	5	8	16	50
	7	120,226	171,030	1.42	1.90	0.84	1	2	4	5	8	16	45
	8	121,212	202,551	1.67	2.11	0.87	1	2	4	6	10	18	38
LA	1	127,584	305,786	2.40	2.39	0.88	2	3	5	7	10	16	49
	2	127,291	252,480	1.98	2.16	0.88	1	3	5	6	9	15	53
	3	130,003	290,283	2.23	2.61	0.87	2	3	5	7	12	20	59
	4	128,079	238,620	1.86	2.35	0.86	1	3	5	6	11	19	51
	5	126,140	241,127	1.91	2.33	0.87	1	3	5	6	11	19	60
	6	124,162	188,670	1.52	2.05	0.87	1	2	4	5	9	18	48
	7	120,154	165,291	1.38	1.95	0.87	1	2	4	5	9	17	52
	8	121,061	219,788	1.82	2.41	0.88	1	2	4	6	11	22	46
MA	1	127,855	297,459	2.33	2.37	0.88	2	3	5	7	10	16	63
	2	127,557	330,598	2.59	2.45	0.87	2	4	6	7	11	16	58
	3	130,303	354,078	2.72	2.93	0.88	2	4	6	8	13	23	48
	4	128,314	358,883	2.80	3.03	0.89	2	4	6	8	14	23	58
	5	126,419	336,758	2.66	2.87	0.87	2	4	6	8	13	22	68
	6	124,340	305,532	2.46	2.71	0.84	2	3	6	8	12	21	55
	7	120,330	247,268	2.05	2.51	0.87	1	3	5	7	12	20	50
	8	121,272	373,309	3.08	3.27	0.86	2	4	7	9	15	25	60

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

Content	Grade	N	No. of WTR Erasures	Mean	Std	Corr. With ERA	Number of erasures by percentiles						Max
							50	75	90	95	99	99.9	
RD	1	127,630	140,269	1.10	1.36	0.86	1	2	3	4	6	9	28
	2	127,324	107,662	0.85	1.19	0.87	0	1	2	3	5	8	32
	3	130,076	140,438	1.08	1.42	0.83	1	2	3	4	6	11	23
	4	128,105	121,404	0.95	1.38	0.85	1	1	3	3	6	11	34
	5	126,221	124,642	0.99	1.36	0.84	1	1	3	4	6	10	31
	6	124,216	94,336	0.76	1.17	0.84	0	1	2	3	5	9	26
	7	120,226	99,306	0.83	1.22	0.84	0	1	2	3	5	10	22
	8	121,212	126,133	1.04	1.46	0.87	1	2	3	4	6	12	30
LA	1	127,584	199,196	1.56	1.67	0.88	1	2	4	5	7	11	43
	2	127,291	164,891	1.30	1.52	0.88	1	2	3	4	6	10	43
	3	130,003	188,710	1.45	1.86	0.87	1	2	4	5	8	14	27
	4	128,079	147,479	1.15	1.61	0.86	1	2	3	4	7	13	42
	5	126,140	149,228	1.18	1.62	0.87	1	2	3	4	7	13	34
	6	124,162	116,331	0.94	1.40	0.87	0	1	3	4	6	11	25
	7	120,154	102,841	0.86	1.36	0.87	0	1	2	3	6	12	42
	8	121,061	133,431	1.10	1.67	0.88	1	2	3	4	7	16	32
MA	1	127,855	199,458	1.56	1.70	0.88	1	2	4	5	7	11	53
	2	127,557	210,834	1.65	1.71	0.87	1	2	4	5	7	11	52
	3	130,303	228,045	1.75	2.13	0.88	1	3	4	6	10	17	31
	4	128,314	234,234	1.83	2.22	0.89	1	3	4	6	10	17	53
	5	126,419	210,098	1.66	2.04	0.87	1	2	4	5	9	16	34
	6	124,340	170,459	1.37	1.77	0.84	1	2	3	5	8	14	37
	7	120,330	148,840	1.24	1.71	0.87	1	2	3	4	8	14	31
	8	121,272	207,594	1.71	2.08	0.86	1	2	4	6	9	16	43

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
1	1244	438	116	371	68
2	1249	422	98	365	51
3	1250	345	30	307	22
4	1248	346	36	287	12
5	1247	309	32	257	21
6	602	220	43	180	14
7	576	223	39	170	10
8	581	216	52	180	26

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 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
1	1244	953	77%	1244	940	76%	1244	936	75%	1244	758	61%
2	1249	979	78%	1249	990	79%	1249	951	76%	1249	770	62%
3	1250	1007	81%	1249	1030	82%	1249	1056	85%	1250	811	65%
4	1248	1025	82%	1248	1034	83%	1248	1091	87%	1248	832	67%
5	1247	1037	83%	1247	1051	84%	1247	1088	87%	1247	872	70%
6	601	441	73%	602	451	75%	600	472	79%	602	345	57%
7	575	441	77%	576	427	74%	576	432	75%	576	323	56%
8	581	424	73%	581	413	71%	581	448	77%	581	326	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
1893	1328	70%	1892	1331	70%	1891	1378	73%	1893	1794	95%

Recommendations for State Board Consideration

February, 2011

	State Action	LEA Action	LEA Student Support
Severe Concern (25% or more of classrooms flagged)	State Monitors during Spring 2011 CRCT	1) Rotate teachers during 2011 CRCT 2) Remove Test Administrators in 2011 whose classrooms were flagged in 2010 across multiple subjects at ≥ 4 SD	Use data files provided by OSA to: a) support independent LEA investigations; and b) Identify students adversely affected (if applicable) who may need supports or interventions.
Moderate Concern (11%-24% of classrooms flagged)	State Monitors during Spring 2011 CRCT	1) Rotate teachers during 2011 CRCT 2) Remove Test Administrators in 2011 whose classrooms were flagged in 2010 across multiple subjects at ≥ 4 SD	Use data files provided by OSA to: a) support independent LEA investigations; and b) Identify students adversely affected (if applicable) who may need supports or interventions.
Minimal Concern (6%-10% of classrooms flagged)		1) Rotate teachers during 2011 CRCT 2) Remove Test Administrators in 2011 whose classrooms were flagged in 2010 across multiple subjects at ≥ 4 SD	Use data files provided by OSA to offer support services as appropriate based on any concerning irregularities