THE **GOVERNOR'S OFFICE**
of **STUDENT ACHIEVEMENT**

Statewide Analysis: Spring 2009 CRCT Grades 1-8
Presented to the State Board of Education on February 10, 2010

Rationale for this Investigation

The Governor's Office of Student Achievement is charged with auditing and inspecting schools and Local Education Agencies (O.C.G.A. § 20-12-24). A comprehensive analysis of the 2009 Spring CRCT data 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 tampering 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 results indicate an accurate representation of students' knowledge.

Purpose of the Criterion Referenced Competency Test (CRCT)

The CRCT is a standardized assessment administered to students in grades 1-8 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 analysis of erasures on the answer sheets used in the spring 2009 CRCT. The analysis was conducted for grades 1-8, in Reading, Language Arts and Math and was designed to identify whether answers were changed from wrong to right more frequently in any classrooms compared to the rest of the state test population in each grade and subject.

Using a professional grade scanner, CTB scanned the answer documents to determine the total number of erasures and the total number of wrong-to-right (WTR) changes on each answer 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 aggregated these results at the school level.

Erasure Analysis Results

The analysis indicates that some classrooms show an unusually high number of wrong answers changed to right answers on the grades 1-8 Spring 2009 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 89.5%, with a statewide average of 4%.

GOSA placed schools into four categories based on the percentage of classrooms within each school flagged according to CTB's results: Clear of concern; Minimal concern; Moderate concern; and Severe concern.

80% of Georgia's elementary and middle schools fell into the "Clear" category, 10% fell into "Minimal concern," 6% fell into "Moderate concern," and 4% fell into the "Severe concern" category.

Erasure Analysis

Submitted by CTB-McGraw Hill

January 22, 2010

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 2009 CRCT that a) were erased and b) changed from wrong to right answers

Reviewers should note that results should only be used to facilitate identification of systematic problems within individual schools. That is, these types of analyses must be supported by additional, collateral information.

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

$$Mean(m_i) = \mu \tag{1}$$

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

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 $\mu + 3\dfrac{\sigma}{\sqrt{n_i}}$. 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 A.1 reports the state summary for erasure counts. The mean number of erasures across grades and subjects ranged from 1.40 to 3.07 for the 2009 spring CRCT. The erasure count at the 95th percentile point was between 5 and 9. Table A.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 95th percentile point were lower than those obtained from all erasure counts. The mean number of wrong-to-right erasures ranged from 0.81 to 1.91 for the 2009 spring CRCT. The wrong-to-right erasure count at the 95th percentile point was between 3 and 6.

Table 1 summarizes all erasure analyses and wrong-to-right erasure analyses. Table 1 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 2.

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 3.

Table 1. 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	1228	478	191	455	180
2	1230	424	174	392	161
3	1232	368	94	316	80
4	1233	298	68	250	50
5	1235	316	90	252	69
6	591	220	45	154	23
7	567	204	55	148	23
8	566	220	65	156	39

Table 2. 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
1	1226	852	69%	1226	880	72%	1226	845	69%	1226	700	57%
2	1229	909	74%	1229	914	74%	1229	903	73%	1229	771	63%
3	1231	969	79%	1231	994	81%	1231	1014	82%	1231	816	66%
4	1231	1036	84%	1231	1019	83%	1230	1063	86%	1231	884	72%
5	1233	1019	83%	1232	1007	82%	1233	1057	86%	1233	879	71%
6	590	436	74%	589	438	74%	589	484	82%	590	350	59%
7	566	414	73%	566	434	77%	566	449	79%	566	341	60%
8	564	401	71%	564	409	73%	564	450	80%	564	330	59%

Table 3. 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% flags across grades	No. of Schools with <1% flags across grades	No. of Schools	No. of Schools with <1% flags across grades	No. of Schools with <1% flags across grades	No. of Schools	No. of Schools with <1% flags across grades	No. of Schools with <1% flags across grades	No. of Schools	No. of Schools with <1% flags across grades	No. of Schools with <1% flags across grades
1857	1191	64%	1857	1220	66%	1857	1260	68%	1857	1594	86%

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.

<u>Appendix</u>

Scanning Operations Documentation

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.

This memorandum is designed to provide some additional documentation in support of a research study that investigated student responses on the Reading, English Language Arts, and Mathematics Spring 2009 CRCT that a) were erased and b) changed from wrong to right. Specifically, it provides information regarding the scanning operations as they pertain to erasure analyses and associated quality control checks.

It should be emphasized that results from erasure analyses performed in 2009 should only be used to identify potential problems within individual classrooms. That is, these types of analyses must be supported by additional investigations and 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.

An erasure analysis involves an analysis of scanner data where additional marks have been made on student answer documents. As part of our quality control check for the 2009 CRCT erasure analyses, we pulled 19487 student papers from 17 schools – and checked whether the wrong-to-right erasures could be confirmed. Over 96% of answer changes were confirmed by raters. The remaining 4% of answer changes were likely due to unintended marks on answer documents made and were distributed throughout the 20K papers examined. This extensive review of the CRCT answer changes indicates that the scanning software employed in the erasure analysis accurately identified students changing wrong to right answers.

Table A.1. State summary for all types of erasure counts

Content	Grade	N_state	ERA_state	Mean_state	Std_state	Min_state	Max_state	P10_state	P25_state	P50_state	P75_state	P90_state	P95_state
RD	1	128,257	221,495	1.73	2.11	0	31	0	0	1	3	4	6
	2	129,365	181,429	1.40	1.87	0	40	0	0	1	2	4	5
	3	128,837	251,810	1.95	2.70	0	44	0	0	1	3	5	7
	4	125,267	212,860	1.70	2.44	0	50	0	0	1	2	4	6
	5	124,798	251,418	2.01	2.71	0	43	0	0	1	3	5	7
	6	120,336	174,097	1.45	2.04	0	49	0	0	1	2	4	5
	7	120,692	168,982	1.40	1.98	0	41	0	0	1	2	4	5
	8	120,859	188,964	1.56	2.12	0	41	0	0	1	2	4	5
LA	1	128,226	306,476	2.39	2.65	0	34	0	1	2	3	6	7
	2	129,328	259,145	2.00	2.37	0	50	0	0	1	3	5	6
	3	128,767	301,406	2.34	2.85	0	55	0	0	2	3	6	8
	4	125,239	276,556	2.21	2.83	0	60	0	0	1	3	5	7
	5	124,744	285,415	2.29	2.85	0	53	0	0	1	3	5	7
	6	120,283	191,496	1.59	2.29	0	56	0	0	1	2	4	6
	7	120,629	173,625	1.44	2.10	0	49	0	0	1	2	4	5
	8	120,724	203,763	1.69	2.35	0	52	0	0	1	2	4	6
MA	1	128,434	321,641	2.50	2.70	0	40	0	1	2	4	6	8
	2	129,540	328,724	2.54	2.72	0	61	0	1	2	4	6	8
	3	129,046	367,994	2.85	3.25	0	57	0	1	2	4	7	9
	4	125,470	368,661	2.94	3.39	0	68	0	1	2	4	7	9
	5	125,019	368,460	2.95	3.48	0	60	0	1	2	4	7	9
	6	120,478	308,007	2.56	2.96	0	67	0	1	2	4	6	8
	7	120,777	249,930	2.07	2.62	0	54	0	0	1	3	5	7
	8	120,953	342,054	2.83	3.33	0	66	0	1	2	4	7	9

Table A.2. State summary for wrong-to-right (WTR) erasure count

Content	Grade	N_state	WTR_state	Mean_state	Std_state	Min_state	Max_state	P10_state	P25_state	P50_state	P75_state	P90_state	P95_state
RD	1	128,257	138,593	1.08	1.49	0	23	0	0	1	2	3	4
	2	129,365	118,288	0.91	1.38	0	28	0	0	0	1	2	3
	3	128,837	151,060	1.17	1.87	0	35	0	0	1	2	3	4
	4	125,267	124,590	0.99	1.64	0	36	0	0	0	1	3	4
	5	124,798	141,789	1.14	1.83	0	37	0	0	1	2	3	4
	6	120,336	98,587	0.82	1.30	0	29	0	0	0	1	2	3
	7	120,692	97,579	0.81	1.27	0	33	0	0	0	1	2	3
	8	120,859	111,596	0.92	1.39	0	35	0	0	1	1	2	3
LA	1	128,226	205,551	1.60	1.95	0	30	0	0	1	2	4	5
	2	129,328	170,610	1.32	1.73	0	43	0	0	1	2	3	4
	3	128,767	194,150	1.51	2.05	0	37	0	0	1	2	4	5
	4	125,239	175,974	1.41	2.03	0	41	0	0	1	2	4	5
	5	124,744	178,967	1.43	2.04	0	34	0	0	1	2	4	5
	6	120,283	115,767	0.96	1.55	0	36	0	0	0	1	3	4
	7	120,629	110,213	0.91	1.48	0	41	0	0	0	1	3	4
	8	120,724	125,520	1.04	1.63	0	41	0	0	1	1	3	4
MA	1	128,434	224,507	1.75	2.05	0	30	0	0	1	3	4	6
	2	129,540	223,945	1.73	2.05	0	44	0	0	1	2	4	5
	3	129,046	241,822	1.87	2.43	0	41	0	0	1	3	5	6
	4	125,470	235,186	1.87	2.44	0	47	0	0	1	3	5	6
	5	125,019	228,146	1.82	2.54	0	45	0	0	1	2	4	6
	6	120,478	171,897	1.43	1.96	0	43	0	0	1	2	4	5
	7	120,777	143,487	1.19	1.77	0	41	0	0	1	2	3	4
	8	120,953	186,083	1.54	2.14	0	45	0	0	1	2	4	5

Formal Recommendations that Require Action by the State Board of Education

Recommendations are intended to eliminate test tampering and help students who were adversely affected by intentional wrongdoing.

	State Action	LEA Corrective Action	LEA Student Support
Severe Concern (25% or more of classrooms flagged)	State Monitors during Spring 2010 CRCT	1) LEA to conduct an investigation 2) Submit results of investigation to OSA 3) Rotate teachers during CRCT.	1) Notify parents 2) Offer student support services as appropriate based on outcomes of LEA investigation
Moderate Concern (11% - 24% of classrooms flagged)	Random Spot Checks by State Monitors	1) LEA to conduct an investigation 2) Submit results of investigation to OSA 3) Rotate teachers during CRCT. 4) Monitor test environment.	1) Notify parents 2) Offer student support services as appropriate based on outcomes of LEA investigation
Minimal Concern (6%-10% of classrooms flagged)		1) LEA to monitor test environment of flagged schools; **OR** 2) LEA to rotate teachers in flagged schools.	1) Offer student support services as appropriate based on any concerning irregularities found.