



COLORADO DEPARTMENT OF EDUCATION

201 East Colfax Avenue • Denver, Colorado 80203-1799
303.866.6600 • www.cde.state.co.us

Robert K. Hammond
Commissioner of Education

Diana Sirko, Ph.D.
Deputy Commissioner

Keith Owen, Ph.D.
Associate Commissioner

State Investigation of Denver Public Schools CSAP Results for 2010 and 2011

After conducting some of its own analyses, Denver Public Schools (DPS) approached the Colorado Department of Education (CDE) regarding acquiring additional analysis on state assessment data. At this time, CDE was in the process of reviewing its own test administration, security and data forensics procedures. Given their synchronous goals, CDE and DPS contracted with the Colorado State Assessment Program's vendor, CTB McGraw Hill, to conduct a state-wide blind analysis, as well as a specific erasure analysis for Denver Public Schools.

The resulting erasure analysis identified two schools with greater than or equal to 70% of their units (i.e., grade/content area combinations) at four standard deviations above the state mean. Given that a multitude of causes can result in a high number of wrong-to-right erasures, no conclusions should ever be made based solely on an erasure analysis. However, it is deemed appropriate that when combined with additional quantitative and qualitative information, determinations regarding test administration may be made. Further investigation was conducted for the two schools identified with very to exceedingly high wrong-to-right erasures: Beach Court Elementary and Hallett Fundamental Academy.

In addition to the erasure analysis, DPS and CDE conducted additional analyses and an investigation firm, Alvarez & Marsal, conducted interviews at the two schools. Based on both the quantitative and qualitative evidence gathered and consultations with DPS, CDE has made determinations regarding the schools' scores. DPS will make appropriate personnel decisions while CDE's certification unit will take action as it deems appropriate regarding certification of individuals involved in any situations which may have resulted in score suppressions.

Beach Court Elementary School

As previously mentioned, Beach Court Elementary met the state established criterion of 70% or more of its units being four or more standard deviations above the state mean in the number of wrong-to-right erasures. High erasure rates indicate that for the group of 3rd, 4th, and 5th grade students, there are an exceedingly high number of wrong-to-right erasures. When these changes are made based on legitimate student response patterns, the relationship between the students' performance on the multiple choice and constructed response patterns should be typical. When these changes have been made outside of standard test administration procedures, aberrant results in terms of the relationship between the multiple choice performance and constructed response performance may exist. The additional analyses conducted by CDE and DPS focused on the difference between the performance on constructed response items and multiple choice items on the CSAP tests.

Table 1: Beach Court Elementary Erasure Analysis vs. State.

Grade	Content	2010			2011		
		State	BC	SD	State	BC	SD
03	Math	0.837	10	48.2	0.956	12.219	50.3
	Reading	0.628	10.289	44.0	0.683	10.408	47.1
	Writing	0.692	8.808	37.2	0.726	8.935	40.7
04	Math	1.265	13.894	50.0	1.278	13.022	53.3
	Reading	1.074	20.5	84.5	0.944	17.673	90.0
	Writing	0.647	12.405	70.6	0.569	10.273	73.4
05	Math	1.414	14.162	50.0	1.372	15.528	57.1
	Reading	1.109	17.761	78.0	0.995	19.535	88.5
	Science	1.352	5.243	15.6	1.4	13.73	46.8
	Writing	0.608	10.283	65.8	0.503	11.326	74.2

For the state analysis, the differences between numbers of correct multiple choice items and constructed response items were computed for three content areas (Reading, Writing and Math)¹ for every grade in every school in the state. Those grade-level differences were then standardized to indicate how discrepant each was from the state mean in standard deviation units (standardized scores are often called z-scores). From a statistical perspective, a discrepancy of four standard deviations would identify only schools with very extreme deviations from the state mean. Such a large deviation would be a rare occurrence.

On the multiple choice-to-constructed response analysis, Beach Court Elementary had standardized scores that were more than four standard deviations greater than the state mean for 2011. Their scores were equally discrepant in 2010. The tables below reveal the z-score discrepancies for 2009, 2010 and 2011 for Beach Court Elementary in Reading, Math and Writing.

Table 2: Beach Court Reading MC CR Difference Z-scores 2009-2011

School	Grade 2009	Count 2009	Z mc-cr diff 2009	Count 2010	Z mc-cr diff 2010	Count 2011	Z mc-cr diff 2011
BEACH COURT ELEMENTARY	3	22	1.95	26	5.71	32	4.68
BEACH COURT ELEMENTARY	4	57	2.83	42	5.33	55	4.11
BEACH COURT ELEMENTARY	5	50	2.65	46	5.80	43	6.86

¹ The CSAP Science assessment was not included in these analyses because the assessment is not given at all grades and the results do not impact the Colorado Growth Model.

Table 3: Beach Court Math MC CR Difference Z-scores 2009-2011

School	Grade 2009	Count 2009	Z mc cr diff 2009	Count 2010	Z mc cr diff 2010	Count 2011	Z mc cr 2011
BEACH COURT ELEMENTARY	3	38	2.22	50	1.94	61	2.11
BEACH COURT ELEMENTARY	4	59	1.91	42	3.33	56	3.46
BEACH COURT ELEMENTARY	5	50	2.48	46	2.82	43	3.44

Table 4: Beach Court Writing MC CR Difference Z-scores 2009-2011

School	Grade 2009	Count 2009	Z mc cr diff 2009	Count 2010	Z mc cr diff 2010	Count 2011	Z mc cr 2011
BEACH COURT ELEMENTARY	3	22	0.00	26	0.97	32	-0.12
BEACH COURT ELEMENTARY	4	57	1.60	42	2.90	55	2.34
BEACH COURT ELEMENTARY	5	50	1.46	46	2.51	43	2.86

DPS also had additional indicators of aberrant results. Two of their analyses are especially relevant. The first analysis was a correlation between students' multiple choice and constructed response performance. Beach Court Elementary was in the bottom 10% of all DPS schools in terms of the relationship. The second analysis determined the percentile rank difference of student performance on the multiple choice items versus the constructed response items. Students with a 20 percentage point difference were flagged. The percentage of Beach Court students flagged exceeded three standard deviations from the mean of all schools in DPS.

The external investigation conducted by Alvarez & Marsal provided qualitative information revealing the following:

- Standardized test administration and/or test security procedures were violated. Test materials were not maintained in the designated secure test storage areas. Rather, the test materials were removed to the principal's office.
- A security breach occurred. The removal of the test materials from the designated secure test storage area to the principal's office is a security breach. No legitimate rationale for this change in location after testing was provided.

- The security breach was not contained in a timely manner. The test materials were kept in the principal's office for about a week before being returned to the district office.
- The violations of test security procedures and the resulting security breach provided opportunity for student test answers to be altered. Results could have been altered in such a way that they did not reflect student independent performance.

Determinations and recommendations regarding Beach Court Elementary: The qualitative information gathered from the interviews revealed a significant security breach, for which no legitimate explanation was provided. The test materials were not held in the designated secure location and were not within the SAL's control following testing. The principal had sole access to the materials for an extended period of time prior to their return to the district office. This evidence combined with the data indicators warrant the conclusion that the Beach Court Elementary scores may not be maintained. Multiple sources of evidence indicate that the significant testing violations occurred at the principal level. The serious, intentional, repeated breaches warrant the suppression of all scores for all grades and content areas from 2010 and 2011 from future growth calculations. In addition, the entire school should participate in intensive training regarding test security procedures and reporting of violations. Lastly, increased monitoring and decreased access to test materials should be considered for Beach Court for future test administrations.

Hallett Fundamental Academy

As previously mentioned, Hallett Fundamental Elementary met the state established criterion of 70% or more of its units being four or more standard deviations above the state mean in the number of wrong-to-right erasures. High erasure rates indicate that for the group of 3rd, 4th, and 5th grade students, there are a very high number of wrong-to-right erasures. When these changes are made based on legitimate student response patterns, the relationship between the students' performance on the multiple choice and constructed response patterns should be typical. When these changes have been made outside of standard test administration procedures, aberrant results in terms of the relationship between the multiple choice performance and constructed response performance may exist. The additional analyses conducted by CDE and DPS focused on the difference between the performance on constructed response item.

Table 5: Hallett Erasure Analysis Compared to State

Grade	Content	2010			2011		
		State	Hallett	SD	State	Hallett	SD
03	Math	0.837	1.396	3.2	0.956	2.25	6.8
	Reading	0.628	0.98	2.2	0.683	2.318	9.3
	Writing	0.692	0.694	0.0	0.726	1.727	5.9
04	Math	1.265	3.476	9.6	1.278	3.021	8.1
	Reading	1.074	2.884	8.0	0.944	3.962	15.8
	Writing	0.647	1.568	5.7	0.569	1.865	9.5
05	Math	1.414	5.5	15.8	1.372	5.07	16.3
	Reading	1.109	4.786	16.5	0.995	3.064	10.3
	Science	1.352	1.429	0.3	1.4	1.419	0.1
	Writing	0.608	3.5	18.8	0.503	2.745	16.1

For the state analysis, the differences between numbers of correct multiple choice items and constructed response items were computed for three content areas (Reading, Writing and Math) for every grade in every school in the state. Those grade-level differences were then standardized to indicate how discrepant each was from the state mean in standard deviation units (standardized scores are often called z-scores). From a statistical perspective, a discrepancy of four standard deviations would identify only schools with very extreme deviations from the state mean. Such a large deviation would be a rare occurrence.

Hallett Fundamental Academy **did not** have any standardized scores that were more than four standard deviations greater than the state mean for 2009, 2010 or 2011, when multiple choice to constructed response rates were compared. In fact, the majority of the standardized scores were within one standard deviation of the state mean. The tables below reveal the z-score discrepancies for 2009, 2010 and 2011 for Hallett Fundamental Academy in Reading, Math and Writing.

Table 6: Hallett Fundamental Academy Read MC CR Difference Z-scores 2009-2011

School	Grade	Count 2009	Z mc-cr diff 2009	Count 2010	Z mc-cr diff 2010	Count 2011	Z mc-cr diff 2011
HALLETT FUNDAMENTAL ACADEMY	3	37	.68	49	-.18	44	.16
HALLETT FUNDAMENTAL ACADEMY	4	40	-.16	44	-.90	52	-.41
HALLETT FUNDAMENTAL ACADEMY	5	41	-2.94	42	.26	47	-.03

Table 7: Hallett Fundamental Academy Math MC CR Difference Z-scores 2009-2011

School	Grade	Count 2009	Z mc-cr diff 2009	Count 2010	Z mc-cr diff 2010	Count 2011	Z mc-cr diff 2011
HALLETT FUNDAMENTAL ACADEMY	3	37	1.35	49	0.84	44	0.77
HALLETT FUNDAMENTAL ACADEMY	4	40	0.39	44	1.75	51	1.38
HALLETT FUNDAMENTAL ACADEMY	5	41	0.44	42	1.38	47	1.06

Table 8: Hallett Fundamental Academy Math MC CR Difference Z-scores 2009-2011

School	Grade	Count 2009	Z mc-cr diff 2009	Count10	Z mc-cr diff 2010	Count11	Z mc-cr diff 2011
HALLETT FUNDAMENTAL ACADEMY	3	37	-1.80	49	-1.22	44	-1.88
HALLETT FUNDAMENTAL ACADEMY	4	40	-.25	44	.28	52	.18
HALLETT FUNDAMENTAL ACADEMY	5	41	-.62	42	-.16	47	.65

DPS also had additional indicators of aberrant results. Two of their analyses are especially relevant. The first analysis was a correlation between students' multiple choice and constructed response performance. DPS compared the percentage correct on multiple choice to percentage correct on constructed response, with the expectation that those percentages should be highly correlated. Hallett Fundamental Academy was not in the bottom 10% of all DPS schools in terms of the relationship. The second analysis determined the percentile rank difference of student performance on the multiple choice items versus the constructed response items. Students with a 20 percentage point difference were flagged. The percentage of Hallett Fundamental Academy students flagged did not exceed three standard deviations from the mean of all schools in DPS.

The external investigation conducted by Alvarez & Marsal provided qualitative information revealing the following:

- Although there is evidence that technical violation regarding standardized test administration and/or test security procedures may have occurred, the interviews did not reveal major violations.
- Again, although there is evidence that technical violations regarding standardized test administration and/or test security procedures may have occurred, the interviews did not reveal any security breached.
- The interviews produced a possible explanation for the high wrong-to-right erasures. Although not consistent across all interviews, more than one interviewee mentioned a test taking strategy teaching students to cross out known wrong answers. Such a practice would result in high rates of wrong-to-right erasures as the students only mark known wrong answers. The interviews did not reveal conclusive evidence regarding post testing opportunities to alter test results.

Determinations and recommendations for Hallett Fundamental Academy: Although the erasure analysis indicated aberrant results, the additional CDE and DPS data analyses did not indicate a

pattern of results consistent with post-administration manipulation of test results through the changing of multiple choice answers. Further, the qualitative information gathered from the interviews revealed technical violations such as uncertain chain of custody for secure test materials and irregular handling of secure test materials. The combined quantitative and qualitative evidence does not warrant the suppression of any of the Hallett Fundamental Academy's test scores. We found no evidence that the principal engaged in any wrong doing. The technical violations do indicate a need for additional training in standard test administration and test security procedures. To ensure that scores are viewed as credible, all test administration and test security procedures must be followed. Increased monitoring during test administration should be considered for Hallett Fundamental Academy.

Isolated Incident Investigation:

Determinations and recommendations: After conducting a series of data reviews and receiving erasure analyses for schools with statistical irregularities at a single grade level or test administration, Denver Public Schools requested the assistance of the Colorado Department of Education to further examine 2010 and 2011 CSAP data in such limited instances. Alvarez & Marsal investigators examined the data and procedures and interviewed staff at three specific schools. CDE staff carefully reviewed the Alvarez & Marsal interview reports regarding staff and student interviews conducted at these schools.

In reports from the three schools, the conclusions of the interviewers were that there was little concrete information indicative of post-administration manipulation of data. The Colorado Department of Education does not analyze teacher level data at this time, and is building teacher-student data links at the state level for future use. CDE staff did examine the data for all the individual students in the schools and grades in question, and found no schoolwide wrong doing and no evidence to support individual classroom post-administration manipulation of data. The Colorado Department of Education recommends additional training on security but does not recommend any further action at these schools.