

**Washington, D.C.
Comprehensive Assessment
System (DC CAS)**

**2011 Wrong-to-Right (WTR)
Answer Changes Report**

July 15, 2011



2011 DC CAS Wrong-to-Right Answer Changes Report

With the high-stakes nature of large-scale statewide assessment programs, there can be situations in which student responses, and hence their scores, may not be a true representation of students' 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 being altered after testing. To maintain the integrity of DC CAS and the validity of the results, it is important that any such instances be discovered.

This study examines incorrect student responses to multiple choice items on the Spring 2011 DC CAS Mathematics, Reading, Science and Biology that were erased and changed to correct responses. We refer to these answer changes as wrong-to-right (WTR) answer changes. Inordinate numbers of WTR answer changes in a specifically identifiable testing administration group may indicate inappropriate intervention on students' answer documents by an educator.

We emphasize that the results from this study may be used in conjunction with other information to investigate whether inappropriate interventions may have taken place. Inordinate WTR answer change rates, by themselves, may simply be coincidental and do not necessarily indicate inappropriate behavior.

Method

The basis for these analyses is to count instances where an incorrect answer choice was erased and replaced by the correct answer choice. The data files for these analyses were generated from the answer document scanned data files (i.e., Winscore files). These files contain information about dark bubbles that unambiguously indicate a student's final answer choice and any faint bubbles that clearly indicate an answer choice that was erased. All analyses are conducted for each test administration group, where each test administration group is indicated in the data files by a teacher's name and school name. Schools determine how teachers are assigned to test administration groups. The test taker and item response validation rules used to retain students and responses to items for the WTR answer change analysis are the same rules for calibration and other analyses reported in the 2011 Operational Technical Report.

The WTR answer change analyses include a statistical test of the null hypothesis (H_0) that the mean WTR answer changes for a test administration group constitutes a random sample from the state distribution of WTR answer changes. The hypothesis is tested against the right-sided alternative (H_1), that the mean number is too high to be explained by random sampling. Test administration groups for which H_0 is rejected are flagged. The central limit theorem in statistics tells us that the sampling distribution of mean number of WTR answer changes for class i (m_i) is asymptotically normal with mean and standard deviation

$$\text{Mean}(m_i) = \mu \quad (1)$$

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

where n_i denotes the size of test administration group i and m_i denotes the mean number of WTR answer changes in test administration group i , respectively. In addition, μ and σ denote the mean and the standard deviation of the distribution of the number of WTR answer changes of the population of individual students with valid DC CAS scores in 2011.

Test administration groups were flagged if their m_i was larger than

$$\mu + 4 \frac{\sigma}{\sqrt{n_i}} \quad (3)$$

This flagging criterion of 4 standard deviations above the state mean provides a statistically conservative test. The standard normal table shows that under random sampling the asymptotic probability of observing a sample mean more than four standard deviations above the population mean is around 0.0001, or one in ten thousand. Even with this conservative test, rejection of H_0 tells us only that the observed mean number of answer changes in a test administration group is unlikely to be the result of random sampling, and nothing beyond that with any certainty.

The denominator in the formula for the state standard deviation (equation 2) indicates that the flagging criterion for each test administration group is adjusted for the number of test takers in a test administration group with valid test scores. For example, if the state mean and SD of wrong to right answer changes are 1.73 and 2.11, respectively, the flagging criterion for a class size of 20 is adjusted to 3.62. For example,

$$1.73 + 4 \frac{2.11}{\sqrt{20}} = 3.62$$

This adjustment ensures that the flagging criterion is equally stringent for classes with considerably different numbers of test takers.

Results

A summary of the numbers of schools and test administration groups included in these analyses for each grade and content area is provided in Table 1.

A state summary of WTR answer changes of individual students (i.e. not test administration groups) is provided in Table 2. The summary includes means and standard deviations as well as the minimum, maximum, and some percentiles of WTR answer changes. The mean number of answer changes across grades and content areas ranges from 0.52 to 2.15, or approximately 1 to 2 WTR answer changes per student answer sheet on average (see the column headed "State Mean No. of WTR changes per student").

Table 3 provides the total number and percentage of test administration groups flagged for inordinate WTR answer changes for each grade and content area. Of the 5,089 test administration groups across grade and content areas, 128 groups (2.5%) were flagged for inordinate WTR answer change rates.

Table 4 indicates the numbers of test administration groups in each grade and content area that were flagged for inordinate WTR answer change rates for the 2008 through 2011 test administrations.

In addition, an Excel file that accompanies this report displays data on WTR answer changes rates for all test administration groups in all grades and content areas in 2011 (see *Answer_Changes_DataFile_DC CAS_Spring2011.xls*). The "Test Administration Group" tab in the file lists each school and the teacher name associated with each test administration group (see column F in the spreadsheet.). It also includes

- (a) the number of students with valid test administrations in each grade and content area (see column G);
- (b) descriptive statistics for all types of answer changes (columns H and I) and WTR answer changes (columns J-L);
- (c) the state mean and standard deviation for WTR answer changes (see columns M and N);
- (d) the flagging criterion adjusted for the number of students in each testing group, as defined in equation 3 (see column O); and
- (e) a '1' to flag WTR answer change rates that exceed the adjusted flagging criterion (see column P).

**Table 1. Number of Schools and Test Administration Groups by Grade and Content Area:
2011**

Grade	Mathematics		Reading		Science		Biology	
	No. of Schools	No. of Test Administration Groups	No. of Schools	No. of Test Administration Groups	No. of Schools	No. of Test Administration Groups	No. of Schools	No. of Test Administration Groups
03	139	303	139	303	--	--	--	--
04	138	294	138	292	--	--	--	--
05	143	292	143	293	142	285	--	--
06	109	302	109	302	--	--	--	--
07	108	317	108	318	--	--	--	--
08	110	302	107	297	107	281	--	--
10	84	331	87	334	--	--	--	--
HS	--	--	--	--	--	--	73	243

Note: "--" = Not Applicable.

Table 2. State Summary of Wrong-To-Right (WTR) Answer Changes: 2011

Content	Grade	No. of Items	No. of Students	State Total No. of WTR Changes	State Mean No. of WTR Changes per Student	State SD of No. of WTR Changes per Student	State Min of WTR	State Max of WTR	State P25 of WTR	State P50 of WTR	State P75 of WTR	State P90 of WTR	State P95 of WTR	State P99 of WTR	State P99.9 of WTR
MA	03	65	4805	10309	2.15	2.63	0	26	0	1	3	5	7	12	20
MA	04	65	4858	5938	1.22	2.00	0	30	0	1	2	3	4	9	24
MA	05	65	4812	4897	1.02	1.44	0	17	0	1	1	3	4	6	14
MA	06	65	4423	3940	0.89	1.26	0	16	0	1	1	2	3	5	10
MA	07	65	4458	4206	0.94	1.26	0	15	0	1	1	2	3	6	9
MA	08	65	4354	4590	1.05	1.52	0	23	0	1	2	3	4	7	14
MA	10	65	4415	3370	0.76	1.14	0	16	0	0	1	2	3	5	8
RD	03	62	4773	8261	1.73	2.38	0	41	0	1	2	4	6	11	20
RD	04	62	4817	4772	0.99	1.70	0	23	0	0	1	3	4	8	18
RD	05	62	4791	3490	0.73	1.19	0	12	0	0	1	2	3	6	10
RD	06	62	4393	3770	0.86	1.40	0	26	0	0	1	2	3	6	12
RD	07	62	4440	3148	0.71	1.14	0	18	0	0	1	2	3	5	10
RD	08	63	4310	3629	0.84	1.23	0	13	0	0	1	2	3	5	9
RD	10	63	4442	3675	0.83	1.33	0	24	0	0	1	2	3	6	12
SC	05	59	4764	2854	0.60	0.98	0	17	0	0	1	2	3	4	8
SC	08	59	4213	2401	0.57	1.06	0	28	0	0	1	2	2	4	8
BI	HS	59	3760	1948	0.52	0.84	0	6	0	0	1	2	2	3	5

Table 3. Number and Percentage of Test Administration Groups Flagged: 2011

Grade	No. of Test Administration Groups	No. of Test Administration Groups Flagged	Percentage of Test Administration Groups Flagged
Mathematics			
03	303	21	6.9
04	294	10	3.4
05	292	9	3.1
06	302	10	3.3
07	317	8	2.5
08	302	6	2.0
10	331	4	1.2
Reading			
03	303	19	6.3
04	292	14	4.8
05	293	9	3.1
06	302	5	1.7
07	318	4	1.3
08	297	3	1.0
10	334	1	0.3
Science			
05	285	1	0.4
08	281	3	1.1
Biology			
HS	243	1	0.4
Total Across Grade & Content Areas			
Total	5089	128	2.5

Table 4. Numbers of Test Administration Groups Flagged for Inordinate WTR Answer Change Rates: 2008 to 2011

	No. of Test Administration Groups Flagged			
Grade	2011	2010	2009	2008
Mathematics				
03	21	29	27	26
04	10	17	26	19
05	9	14	26	21
06	10	17	14	16
07	8	9	12	13
08	6	9	8	11
10	4	4	5	11
Reading				
03	19	24	26	22
04	14	16	21	23
05	9	11	26	24
06	5	16	18	18
07	4	9	13	9
08	3	9	9	11
10	1	1	7	8
Science				
05	1	9	11	N/A
08	3	7	2	N/A
Biology				
HS	1	4	2	N/A
All Grades and Content Areas				
Total	128	205	253	232

Final Remarks

This study examines incorrect student responses to multiple choice items on the spring 2011 DC CAS Mathematics, Reading, Science, and Biology assessments that were erased and changed to correct responses, or WTR answer changes. We repeat for emphasis that the results from this study may be used in conjunction with other information to investigate whether inappropriate interventions in test administrations or with answer documents may have taken place. Inordinate WTR answer change rates, by themselves, do not necessarily indicate inappropriate interventions.