

**Missouri Assessment Program  
ELA Grades 3-8  
Mathematics Grades 3-8  
Science Grades 5 and 8**

**Spring 2015 Wrong-to-Right (WTR) Answer  
Changes Report**

**Data Recognition Corporation | CTB  
August 4, 2015**

## 2015 MAP 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 Missouri MAP 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 2015 MAP ELA, Mathematics, and Science that were 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 replaced by the correct answer choice. All operational multiple-choice items were included in the analyses. Answer changes were captured during the on-line test administration. For paper-and-pencil, Braille, and large print form, the changes were captured during transcription into the appropriate on-line form. The analyses were conducted for each test administration group, where a test administration group is indicated in the data files by a combination of a district code, school code, school name, and an educator's first and last name within a given content area and grade level.

A basic data cleaning was conducted to remove duplicate cases. Duplicate student records were removed from the analyses. If duplicate items for the same student were encountered and one displayed WTR change then the changed record was retained and considered as WTR change. If identical duplicate items for the same student were encountered, duplicates were excluded from the data analysis.

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 and  $m_i$  denotes the mean number of WTR answer changes for test administration group  $i$ , respectively. In addition,  $\mu$  and  $\sigma$  denote the mean and the standard deviation of the distribution of the number of WTR answer changes of the population of individual students in MAP 2015.

Test administration groups were flagged if their  $m_i$  was larger than  $\mu + 4 \frac{\sigma}{\sqrt{n_i}}$ .

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 WTR answer changes are 1.73 and 2.11, respectively, the flagging criterion for a class size of 20 is adjusted to 3.62 (i.e.,  $1.73 + 4 \frac{2.11}{\sqrt{20}} = 3.62$ ). 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.

Table 2 provides the total number and percentage of test administration groups flagged for WTR answer changes for each grade and content area. Of the 18353 test administration groups for ELA, 117 (0.64%) groups were flagged for WTR answer changes larger than the flagging criterion across all grade levels. For Mathematics, 131 (0.73%) test administration groups out of total 17876 test administration groups were flagged across all test levels. For Science, of the 5070 test administration groups, 27 (0.53%) groups were flagged. Overall, the percentage of test administration groups flagged for WTR answer changes larger than the flagging criterion was small.

A state summary of WTR answer changes of individual students (i.e., not test administration groups) is provided in Table 3. The summary includes means and standard deviations as well as the minimum, maximum, and selected percentiles of WTR answer

changes. The mean number of WTR answer changes across grades for ELA ranges from 1.17 to 1.65, or approximately 1-2 WTR answer changes per student on average (see the column headed State Mean of WTR). For Mathematics tests the mean number of WTR answer changes across grades ranges from 0.32 to 0.79, or up to 1 WRT change per student on average. For Science tests, the mean number of WTR answer changes was 0.98 for Grade 5 and 1.04 for Grade 8, or approximately 1 WRT change per student on average.

The Excel file that accompanies this report shows the numbers of WTR answer changes for test administration groups in districts with the largest numbers of flagged test administrator groups (see MAP2015\_WTR\_5Districts.080315 with most flags.xlsx). The “Test Administration Group” tab in the file lists each school and the teacher indicator (including teacher’s name) associated with each test administration group (see column F in the spreadsheet). It also includes the number of students in each test administration group with:

- (a) number of students in each grade and content area (see column G) associated with this teacher;
- (b) number of students with WTR answer changes (see column H);
- (c) sum and average number of WTR changes for these students (see columns I and J);
- (d) the flagging criterion adjusted for the number of students as defined in equation 2 (see column K);
- (e) the state mean and standard deviation for WTR answer changes (see columns L and M); and
- (f) a “1” to flag WTR answer change rates that do exceed the adjusted flagging criterion (see column N).

The “School Summary” tab in the file is a school level aggregated summary report based on the flags in the “Test Administration Group” tab for the same schools and districts. A data file with all districts and schools is also available and can be provided to DESE upon request.

### **Final Remarks**

This study examines student wrong-to-right response changes to multiple choice items on the spring 2015 MAP ELA, Mathematics, and Science assessments. Note that in many cases the size of the test administration group was smaller than 5 students. We repeat for emphasis that the results from this study should be used in conjunction with other information to investigate whether inappropriate interventions may have taken place, particularly if the number of students in a test administration group is very small. Inordinate WTR answer change rates, by themselves, do not necessarily indicate inappropriate behavior.

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

Grade	English Language Arts		Mathematics		Science	
	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	1127	3884	1127	3879	n/a	n/a
04	1122	3764	1122	3729	n/a	n/a
05	1063	3449	1062	3386	1062	3134
06	780	2538	780	2433	n/a	n/a
07	677	2401	675	2295	n/a	n/a
08	697	2317	694	2154	694	1936

**Table 2. Number and Percentage of Test Administration Groups Flagged**

Grade	No. of Test Administration Groups	No. of Test Administration Groups Flagged	Percentage of Test Administration Groups Flagged
<b>English Language Arts</b>			
03	3884	43	1.1
04	3764	21	0.6
05	3449	15	0.4
06	2538	14	0.6
07	2401	13	0.5
08	2317	11	0.5
<b>Mathematics</b>			
03	3879	41	1.1
04	3729	31	0.8
05	3386	32	0.9
06	2433	12	0.5
07	2295	9	0.4
08	2154	6	0.3
<b>Science</b>			
05	3134	18	0.6
08	1936	9	0.5

**Table 3. State Summary of Wrong-To-Right (WTR) Answer Changes**

Content	Grade	No. of MC items		No. of Students	State Sum of WTR	State Mean of WTR	State SD of WTR	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
		On-line forms	CT forms													
ELA	3	27-32	28	67860	112294	1.65	1.64	0	20	0	1	2	4	5	7	11
ELA	4	20-23	32	66872	86913	1.3	1.36	0	16	0	1	2	3	4	6	9
ELA	5	22-25	27	66279	86798	1.31	1.37	0	16	0	1	2	3	4	6	9
ELA	6	24-25	30	65915	98874	1.5	1.5	0	14	0	1	2	3	4	6	9
ELA	7	19-21	24	65834	76931	1.17	1.28	0	16	0	1	2	3	4	5	8
ELA	8	19-24	32	66344	95464	1.44	1.45	0	16	0	1	2	3	4	6	9
MA	3	4-6	12	67985	21806	0.32	0.58	0	6	0	0	1	1	1	2	3
MA	4	6	13	66973	25885	0.39	0.65	0	6	0	0	1	1	2	3	4
MA	5	11-15	21	66394	52193	0.79	0.98	0	12	0	1	1	2	3	4	6
MA	6	3-7	5	65955	25563	0.39	0.62	0	5	0	0	1	1	2	2	3
MA	7	5-7	6	64887	24562	0.38	0.63	0	5	0	0	1	1	2	2	4
MA	8	8-11	10	52745	25489	0.48	0.72	0	6	0	0	1	1	2	3	4
SC	5	18	18	66357	64735	0.98	1.17	0	18	0	1	1	2	3	5	8
SC	8	15-16	16	66383	69171	1.04	1.14	0	12	0	1	2	3	3	5	7

Note: Number of MC items on Transcribed (CT) forms was often different from the number of items on forms administered on-line.