



2007 WEST VIRGINIA ONLINE WRITING ASSESSMENT SPECIAL STUDY



Developed and published under contract with the West Virginia Department of Education by CTB/McGraw-Hill LLC, a subsidiary of the McGraw-Hill Companies, Inc., 20 Ryan Ranch Road, Monterey, California 93940-5703. Copyright © 2007 by the West Virginia Department of Education. Only State of West Virginia educators and citizens may copy, download and/or print the document, located online at <http://WESTEST.k12.wv.us>, for educational purposes only. Any other use or reproduction of this document, in whole or in part, requires written permission of the West Virginia Department of Education and the publisher.

Table of Contents

Executive Summary	5
CTB's Writing Roadmap.....	5
Overview of the report.....	5
Data sources	5
Comparison of WRM 2.0 and WV Online Writing Assessment.....	6
Use of the Writing Roadmap for Practice.....	8
Student distribution	8
Prompt-type and grade-level distribution	8
Practice and Performance Measures.....	9
Analysis procedures.....	9
Operational performance and number of all practice trials	9
Operational performance and number of on-level practice trials	10
Performance and Practice Measures.....	12
Grade 6 Performance and Grade 7 Practice Measures.....	14
Practice and Performance Measures within 2006 Performance Levels.....	16
Summary and Conclusions	18

List of Tables

Table 1. Case counts for datasets	6
Table 2. Comparison of WRM 2.0 and WV Online Writing Assessment.....	7
Table 3. Practice of students.....	8
Table 4. Distribution of prompt types and levels	8
Table 5. Performance measures by number of all trials.....	9
Table 6. Performance measures by number of on-level trials.....	10
Table 7. Practice measures by operational performance level	12
Table 8. Practice measures by performance level	14
Table 9. Summary scores by number of practice trials within 2006 performance levels	16

List of Figures

Figure 1. Summary scores by number of trials 10

Figure 2. Summary scores by on-level trials 11

Figure 3. Number of practice trials by performance level..... 13

Figure 4. Number of all practice trials by performance level on 2006 RLA test 15

Figure 5. Summary scores by number of practice trials within 2006 levels 16

Figure 6. Summary scores for no and 4+ practice trials within levels 17

Executive Summary

The relationship between practice of Writing using WRM 2.0 and performance on the WV Online Writing Assessment is clear and consistent in the data presented in this report. Data from this study indicated that higher performing students are more likely to practice and to practice more than lower performing students. When examined within performance level, those who practice do better than those who do not, and those who practice more do better than those who practice less.

CTB's Writing Roadmap

The Writing Roadmap™ (WRM 2.0) product was developed by CTB as a formative classroom tool for teachers and a complement to the summative assessments. Teachers are able to select one of four types of prompts for a student to write an essay. The online score report provides feedback on several specific dimensions of writing proficiency. This feedback allows teachers to tailor instruction to better meet each student's needs. This report provides results of the examination of the relationship between practice and performance. This study focused on practice with WRM 2.0 as the generalization of writing skills, and did not examine type of prompt or evaluation dimension.

Overview of the report

The first section of this report presents information about the general use of WRM 2.0 from the fall of 2006 to February 23, 2007, immediately prior to the administration of the WV Online Writing Assessment. The second section presents results on the relationship between practice and performance on the assessment, comparing groups by how many practice trials were taken. The third section examines the relationship between performance and practice, comparing the different amounts of practice with WRM 2.0 by students at different performance levels on the operational assessment.

The fourth section presents results, for grade 7 only, on amounts of practice by performance levels that were based on the preceding year's comprehensive Reading and Language Arts section of the WESTEST. Students are not tested in grade 9 in WESTEST; therefore, students in grade 10 do not have test scores from the preceding year. Section five presents results of the effect of practice on the operational assessment, within the performance levels based on the preceding year's RLA test scores. The final section presents a summary and conclusions.

Data sources

Table 1 presents case counts for the datasets used in the analyses. The Online Writing Assessment data comprised 20,311 grade 7 students and 19,327 grade 10 students. All students in the operational data had valid WVEIS numbers. The practice dataset contained information on 61,832 trials. Not all students used valid WVEIS numbers, by which the data were merged. The dataset for the Grade 6 2006 RLA WESTEST

contained 20,782 cases, of which, 19,027 were matched with this year's grade 7 online assessment.

Table 1. Case counts for datasets

Practice Data	Grade 7	Grade 10
Practices	33,874	27,909
Matched practices	33,136	26,836
Students	12,450	10,085
On-level only	7,366	6,650
Merged Data	Grade 7	Grade 10
Online test	20,311	19,327
Analyses	20,311	19,327
2006 RLA	20,782	---
RLA matched cases	19,027	---

Comparison of WRM 2.0 and WV Online Writing Assessment

It is important to note several differences between WRM 2.0 and the WV Online Writing Assessment (Table 2). The WRM 2.0 rates essays on 6 dimensions, whereas the WV Online Writing Assessment rates 5 dimensions; the WRM 2.0 assigns a holistic score which is independent of the 6 dimensions, whereas the summative score in the WV Online Writing Assessment is the sum of the 5 ratings. Though these are not compared in this report, the WRM 2.0 holistic score is on a scale of 0-6, with one decimal place, the WV Online Writing Assessment summary score is 0-30 in integers.

Additionally, for the WV Online Writing assessment, one of four types of on-level prompts is randomly assigned. For practice with the WRM 2.0, teachers could assign any of 34 prompts, also of the four types, but from four different grade groupings (grade 3, grades 4-6, 7-9, 10-12). Scores for these are all at their respective grade levels. Also, teachers could assign more prompts to some students than others, just as some teachers could assign more students to practice than other teachers.

Table 2. Comparison of WRM 2.0 and WV Online Writing Assessment

	WRM 2.0	WV Online Writing
Prompt types	Descriptive Informative Narrative Persuasive	Descriptive Expository Narrative Persuasive
Trait ratings	Content Organization Voice Word Choice Fluency Conventions	Organization Development Sentence structure Word choice Mechanics
Summative Score	Holistic score, independent of 6 trait ratings	Sum of 5 trait ratings
Scoring range	0.0 – 6.0	0 – 30

Use of the Writing Roadmap for Practice

Student distribution

Over half of students had at least one practice trial with WRM 2.0, about the same rates of participation as last year. Table 3 indicates this percent was higher for grade 7 than grade 10 students. For those that did practice, they did so about 2.7 times on average, an increase from last year from 1.8 to 2.7 in grade 7 and, in grade 10, from 2.2 to 2.7. Of the students that did practice, the proportion who did all of that practice with on-grade level prompts only was slightly smaller in grade 7, almost 60%, whereas 2/3 of those in grade 10 practiced with only on-level prompts.

Table 3. Practice of students

	Grade 7	Grade 10
N of students	20,311	19,327
Percent practiced	61%	52%
Ave N of trials	1.7	1.4
N of students practiced	12,450	10,085
Ave N of trials	2.7	2.7
Percent all on-level	59%	66%

Prompt-type and grade-level distribution

Whereas the distribution of prompt type in the operational assessment is equal, the assignment of prompt types by teachers was not uniform. Table 3 indicates that, in grade 7, expository prompts, on which students do better on the WV Online Writing Assessment, were assigned the most, whereas the more difficult narrative prompts were assigned much less often. In grade 10, the more challenging persuasive prompts were assigned the least frequently. As noted in the introduction, WRM 2.0 provides prompt types for four grade-groups. Table 4 indicates that 70 and 76% of the assigned prompts were on-level, for grades 7 and 10, respectively.

Table 4. Distribution of prompt types and levels

Prompt type	Grade 7	Grade 10
Descriptive	24%	27%
Expository	32%	29%
Narrative	23%	22%
Persuasive	21%	22%
Prompt level		
3	3%	3%
4-6	22%	7%
7-9	70%	14%
10-12	5%	76%

Practice and Performance Measures

Analysis procedures

For all the subsequent data presented here, a one-way Analysis of Variance was conducted first to determine if the groups being compared differed among themselves. In every case, there were significant differences. As all comparisons involved ordinal groups, the ANOVAs also indicated a significant linear component. Next, a series of post-hoc comparisons were made to determine, more specifically, such things as whether those with one practice did significantly better than those with none, if those with two practices did significantly better than those with one, and so forth. For each group, a confidence interval is set, and the groups are compared to determine where there is overlap (i.e., non-significantly different). No overlap in the confidence intervals means that the comparison groups differ significantly. Scheffe's method (1959) was selected because of the unequal sizes of the comparison groups.

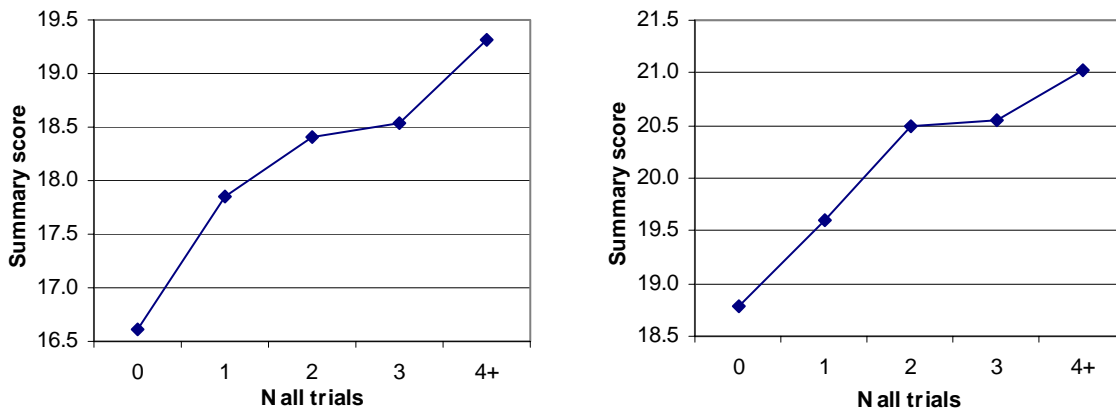
Operational performance and number of all practice trials

Because students were not randomly assigned to the number of prompts they took (nor the type, level, or date), we cannot assume that those who took varying numbers of trials were similar, therefore, students were aggregated into common practice levels (e.g., all those with 3 trials). Table 5 presents the average summary scores and proportion at Mastery, by the number of practice trials taken at any level, with those of 4 or more trials collapsed. Figure 1 displays this data for summary scores. For grades 7 and 10, the number of students were 20,311 and 19,327, respectively.

Table 5. Performance measures by number of all trials

Grade 7					
N all trials	N	Freq	Summary	SD	Mastery
0	7861	39%	16.6	5.1	69%
1	4250	21%	17.9	4.7	78%
2	3463	17%	18.4	4.6	81%
3	2054	10%	18.5	4.6	82%
4+	2683	13%	19.3	4.8	85%
Grade 10					
N all trials	N	Freq	Summary	SD	Mastery
0	9242	48%	18.8	5.4	82%
1	3273	17%	19.6	4.8	89%
2	2537	13%	20.5	4.8	91%
3	2081	11%	20.6	4.6	92%
4+	2194	11%	21.0	4.5	94%

Figure 1. Summary scores by number of trials for grades 7 (left) and 10 (right)



Post-hoc comparisons indicated that each group differed significantly from the others, except trials 2 and 3, which did not differ from each other, in both grades for summary scores.

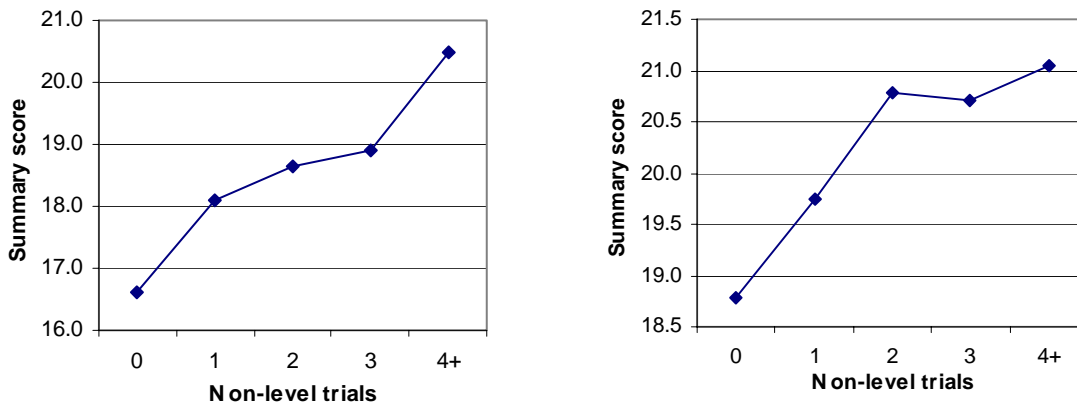
Operational performance and number of on-level practice trials

Table 6 and Figure 2 present performance results, summary scores, and percent at Mastery, for number of practice trials for those whose practices were all on-level, compared to those who had none. For these analyses, the number of students for grades 7 and 10 were 15,227 and 15,892, respectively.

Table 6. Performance measures by number of on-level trials

Grade 7						
N on-level trials	N	Freq	Summary	SD	Mastery	
0	7861	52%	16.6	5.1	69%	
1	3143	21%	18.1	4.7	80%	
2	2038	13%	18.7	4.7	82%	
3	1129	7%	18.9	4.5	83%	
4+	1056	7%	20.5	4.7	90%	
Grade 10						
N on-level trials	N	Freq	Summary	SD	Mastery	
0	9242	58%	18.8	5.4	82%	
1	2351	15%	19.8	4.8	89%	
2	1801	11%	20.8	4.7	92%	
3	1255	8%	20.7	4.6	93%	
4+	1243	8%	21.1	4.5	94%	

Figure 2. Summary scores by on-level trials for grades 7 (left) and 10 (right)



Post-hoc comparisons indicated, for grade 7, that all contrasts were significant except the pairs, 1-2 trials and 2-3 trials. In grade 10, one practice was significantly better than none and two was significantly better than one, but the differences among 2, 3 and 4 or more on-level trials were not significantly different.

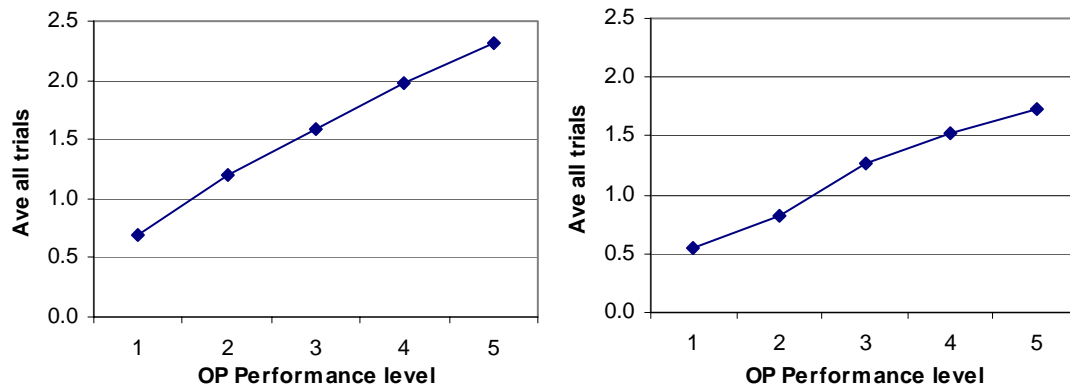
Performance and Practice Measures

This section presents differences in practice measures by those at different performance levels. All students received a WV Online Writing Assessment score in spring 2007 and were, thus, in five performance categories. Table 7 presents evidence that those at higher performance levels were more likely to practice and practiced on more trials. Figure 3 displays the results for number of practice trials. These analyses involved 20,311 and 19,327 students in grades 7 and 10, respectively.

Table 7. Practice measures by operational performance level

Grade 7						
OP Performance level	N	Freq	% practiced	Ave all trials	Ave on-level	
Novice	567	3%	32%	0.7	0.4	
Partial mastery	4279	21%	51%	1.2	0.8	
Mastery	11017	54%	63%	1.6	1.1	
Above mastery	3749	18%	70%	2.0	1.5	
Distinguished	699	3%	75%	2.3	1.9	
Grade 10						
OP Performance level	N	Freq	% practiced	Ave all trials	Ave on-level	
Novice	464	2%	26%	0.5	0.4	
Partial mastery	2055	11%	38%	0.8	0.5	
Mastery	9666	50%	52%	1.3	0.8	
Above mastery	5572	29%	58%	1.5	1.0	
Distinguished	1570	8%	62%	1.7	1.1	

Figure 3. Number of practice trials by performance level for grades 7 (left) and 10 (right)



Post-hoc comparisons indicated that students at each performance level differed significantly from other levels in both the total number of practice trials taken and on-level trials taken, for both grades.

Grade 6 Performance and Grade 7 Practice Measures

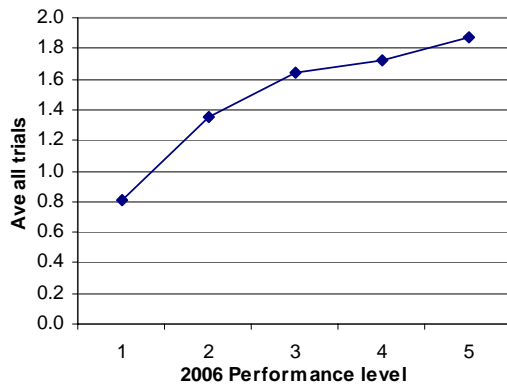
The analyses in the previous section were confounded by the fact that the operational assessment, and thus one's performance level, came after the practice. To avoid this problem, performance levels of grade 7 students from the previous year's Reading and Language Arts (RLA) test were merged with the practice and performance data set. These performance level groups are, thus, identified prior to and independent of the writing assessment. The 2006 WV Online Writing Assessment Technical Report provides evidence that the RLA test is highly correlated (~ 0.60) with the writing assessment. The same analyses could not be performed with grade 10 students, as there is no grade 9 testing in WESTEST.

Similar results as the previous section were found, that is, higher performers were more likely to practice and practiced more, both on- and all-level (Table 8). However, the differences between levels are greatest at the low end of performance and decline between higher levels, as shown in Figure 4. A total of 19,027 students were used in these analyses.

Table 8. Practice measures by performance level on 2006 RLA test

Grade 7					
2006 Performance level	N	Freq	% practiced	Ave all trials	Ave on-level
Novice	727	4%	39%	0.8	0.6
Partial mastery	2531	13%	53%	1.4	1.0
Mastery	8571	45%	64%	1.6	1.1
Above mastery	5235	28%	67%	1.7	1.2
Distinguished	1963	10%	67%	1.9	1.4

Figure 4. Number of all practice trials by performance level on 2006 RLA test



The above comparisons indicated significant differences among all performance groups, except the two pairs, levels 3-4 and 4-5, on both total number of trials taken and on-level trials taken.

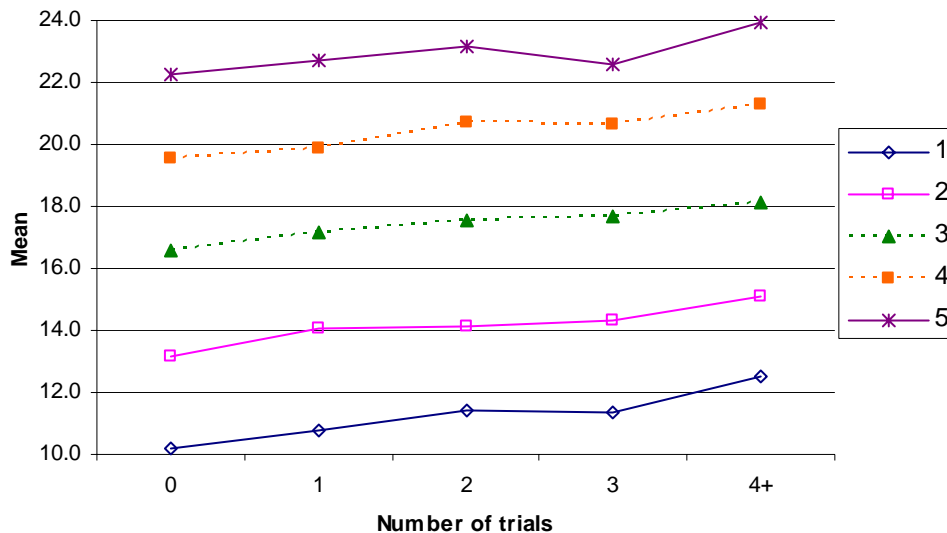
Practice and Performance Measures within 2006 Performance Levels

The results from the previous section make it clear that those who are at higher performance levels tend to practice more. Therefore, practice was examined within each of the RLA performance groups. Table 9 presents summary scores, within level, for those with different numbers of practice trials. In general, within each of the five RLA performance groups, those who practiced more did better on the 2007 Online Writing Assessment. Figure 5 shows the consistent improvements on the Online Writing Assessment within each RLA performance level. Again, 19,027 students were used in these analyses.

Table 9. Summary scores by number of practice trials within 2006 performance levels

Summary score	Number of practice trials					
	2006 Performance level	0	1	2	3	4+
Novice		10.2	10.8	11.4	11.4	12.5
Partial mastery		13.1	14.0	14.1	14.3	15.1
Mastery		16.6	17.1	17.5	17.7	18.1
Above mastery		19.6	19.9	20.7	20.7	21.3
Distinguished		22.2	22.7	23.1	22.6	23.9

Figure 5. Summary scores by number of practice trials within 2006 performance levels

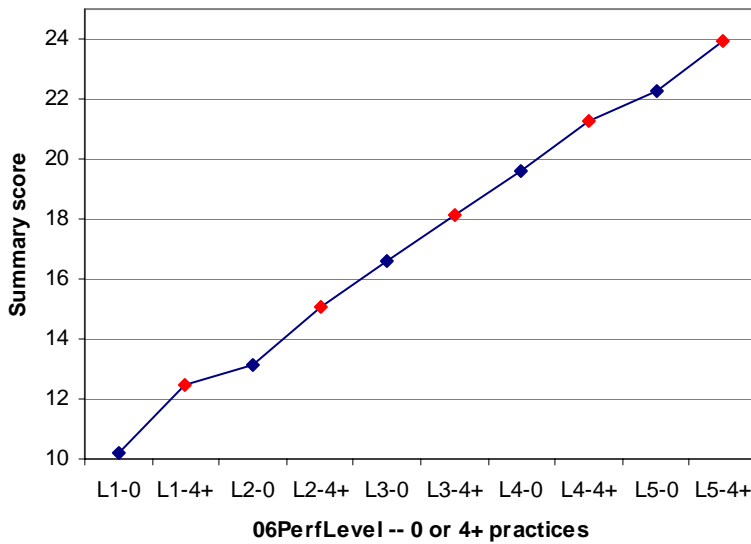


Within each performance level, a significant linear trend was found, though not all adjacent pairs differed significantly.

Performance categories, based on 2006 RLA testing, differed by an average of 3 points on their Online Writing scores. Within each of these categories, repeated practice (4+) showed improvements of a group average of almost 2 points, over 60% of the gap to the average performance of those at the next highest performance category. Furthermore, these high practice groups in each performance category performed only 1 point behind those at the next level who had no practice.

Figure 6 shows the comparison of those with no practice versus those with 4 or more practices, within each of the 2006 performance categories. The consistency of improvement with practice across performance levels is evident in the near linear increase. A total of 9,589 students are represented here.

Figure 6. Summary scores for no (blue) and 4+ (red) practice trials within levels



Summary and Conclusions

There is strong and consistent evidence of a relationship between practice with the Writing Roadmap 2.0 and performance on the WV Online Writing Assessment. Although it is clear that those at higher performance levels were more likely to practice and practiced with more trials than those at lower levels, there is convincing evidence that practice improves performance for all levels of students. When examined within performance levels, those who practiced did better than those who did not, and those who practiced more did better than those who practiced less.

As a general pattern, the largest differences are noted between those who had no practice and those who had one practice, which would include both a familiarity effect with the technology and a learning component. From one to four or more practice trials, gains were consistently positive, which demonstrates a learning effect.

The effects of multiple practices, for all levels of student performance, demonstrates that those who practice frequently will improve their performance on the Online Writing Assessment to significantly above what would be expected if they had no practice. Those who practiced frequently performed almost as well as those at the next level who did not practice.