

2010

Research Brief

Impact of Writing Roadmap 2.0 on WESTEST 2 Online Writing Assessment Scores

This is an internal research document prepared by the WVDE Office of
Assessment, Accountability and Research (OAAR)

Larry White – AD/Psychometrician/Lead Researcher
Nate Hixson – Research and Evaluation Coordinator
Juan D’Brot – Research and Evaluation Coordinator
Jason Perdue – Online Technology/Assessment Coordinator
w/ Sandra Foster and Vaughn Rhudy – OAAR



Contents

INTRODUCTION AND PURPOSE	3
Research Question.....	3
DESIGN AND METHODOLOGY.....	4
AGGREGATED FINDINGS FOR ALL GRADES.....	5
Average Number of Writing Roadmap 2.0 Assignments Completed (All Grades).....	5
Average WESTEST 2 Online Writing Assessment Scores (All Grades).....	5
Significance Testing (All Grades).....	6
FINDINGS BY GRADE LEVEL.....	7
Average Number of Writing Roadmap 2.0 Assignments Completed (By Grade)	7
Average WESTEST 2 Online Writing Assessment Scores (By Grade)	7
Significance Testing (By Grade).....	10
REPLICATION WITH SMALLER RANDOM SAMPLES.....	11
CONCLUSION.....	12
RECOMMENDATIONS	12
REFERENCES.....	13
APPENDIX A: REPLICATION ACROSS GRADE LEVELS.....	14
APPENDIX B: REPLICATION WITHIN GRADES.....	18

INTRODUCTION AND PURPOSE

Beginning with the field test in 2008, *WESTEST 2 Online Writing* became a component of the reading/language arts portion of WESTEST 2. This criterion-referenced test is designed to measure the writing ability of West Virginia students and is administered to all students in grades 3-11 who are not required to take the West Virginia Alternate Performance Task Assessment (APTA).

WESTEST 2 Online Writing assesses student writing skills based on Standard 2: Writing of the 21st Century Reading and English Language Arts Content Standards and Objectives for West Virginia Schools. The purpose of the WESTEST 2 Online Writing assessment is to inform instruction in the important area of written communication and to assist teachers in improving the writing skills of West Virginia students to meet national educational standards, higher education goals and global economic demands.

On a secure Web site, students use a word processing program to enter their responses to assigned passages and accompanying prompts based on four genres: descriptive, narrative, informative and persuasive. These passages and prompts emphasize 21st century skills, including real world applications, critical thinking and problem solving. Student responses are computer scored, using the West Virginia Writing Rubrics. A score of 1 to 6 is assigned in the analytic traits of Organization, Development, Sentence Structure, Word Choice/Grammar Usage and Mechanics.

To help prepare students for *WESTEST 2 Online Writing*, the West Virginia Department of Education offers Writing Roadmap 2.0, a Web-based tool which provides students with the opportunity for unlimited practice sessions for writing essays on a variety of prompts. Writing Roadmap 2.0 was available for all West Virginia public school students in grades 3-12 during the 2008-2009 school year.

In November 2009, the West Virginia Department of Education Office of Assessment, Accountability and Research (WVDE: OAAR) conducted an internal research study to determine the impact of Writing Roadmap 2.0 on West Virginia students' 2009 WESTEST 2 Online Writing assessment scores. This research study utilized extant data to address a single research question:

Research Question

RQ: Is there a statistically significant difference between the 2009 WESTEST 2 Online Writing Assessment scores of West Virginia students who used Writing Roadmap 2.0 during the 2008-2009 school year and those who did not?

DESIGN AND METHODOLOGY

A quasi-experimental design was utilized to answer the study research question. The selection of the treatment group was purposeful, containing all nondisabled West Virginia students in grades 4 through 9 and 11¹ who had completed five or more Writing Roadmap 2.0 exercises during the 2008-2009 academic year. A comparison group was then drawn from the population of nondisabled West Virginia students in grades 4 through 9 and 11 who had not used Writing Roadmap 2.0 during the 2008-2009 school year. The students selected as a comparison group were matched to students in the treatment group based first on grade level², then geographic location (i.e., RESA) and socioeconomic status (i.e., Eligibility for Free and Reduced Price Lunch). The final study sample contained 8,430 students in the treatment condition and 8,430 students in the comparison sample.

OAR conducted descriptive analyses (i.e., frequencies) to confirm that the matching process was successful. Tables 1 and 2 present the results and illustrate that both groups were adequately matched on grade level, geographic location and socioeconomic status. The grade level pairs were exactly matched and in no case did the differences between groups exceed five percentage points for RESA or socioeconomic status.

Table 1: Matching on Grade Level

Group	<i>N</i>	4	5	6	7	8	9	11
Exp.	8,430	14.0%	18.4%	16.4%	14.0%	17.3%	12.0%	7.9%
Comparison	8,430	14.0%	18.4%	16.4%	14.0%	17.3%	12.0%	7.9%

Table 2: Matching on Geographic Location and Socioeconomic Status

Group	<i>n</i>	RESA I	RESA II	RESA III	RESA IV	RESA V	RESA VI	RESA VII	RESA VIII	% Low SES
Exp.	8,430	12.7%	16.2%	15.8%	11.6%	10.0%	8.7%	21.6%	3.4%	48.4%
Comparison	8,430	16.5%	18.4%	14.8%	8.1%	9.3%	9.3%	20.3%	3.2%	53.2%

¹ Grades 3 and 10 were excluded from the current study because researchers initially sought to examine students' 2007-2008 standardized reading/language arts performance. Students in grades 3 and 10 during the 2008-2009 academic year did not have data for this measure.

² Both the treatment and comparison populations were stratified by grade level and students were randomly selected from either group in each grade level to ensure that the same number of students were included in both treatment and comparison conditions.

AGGREGATED FINDINGS FOR ALL GRADES

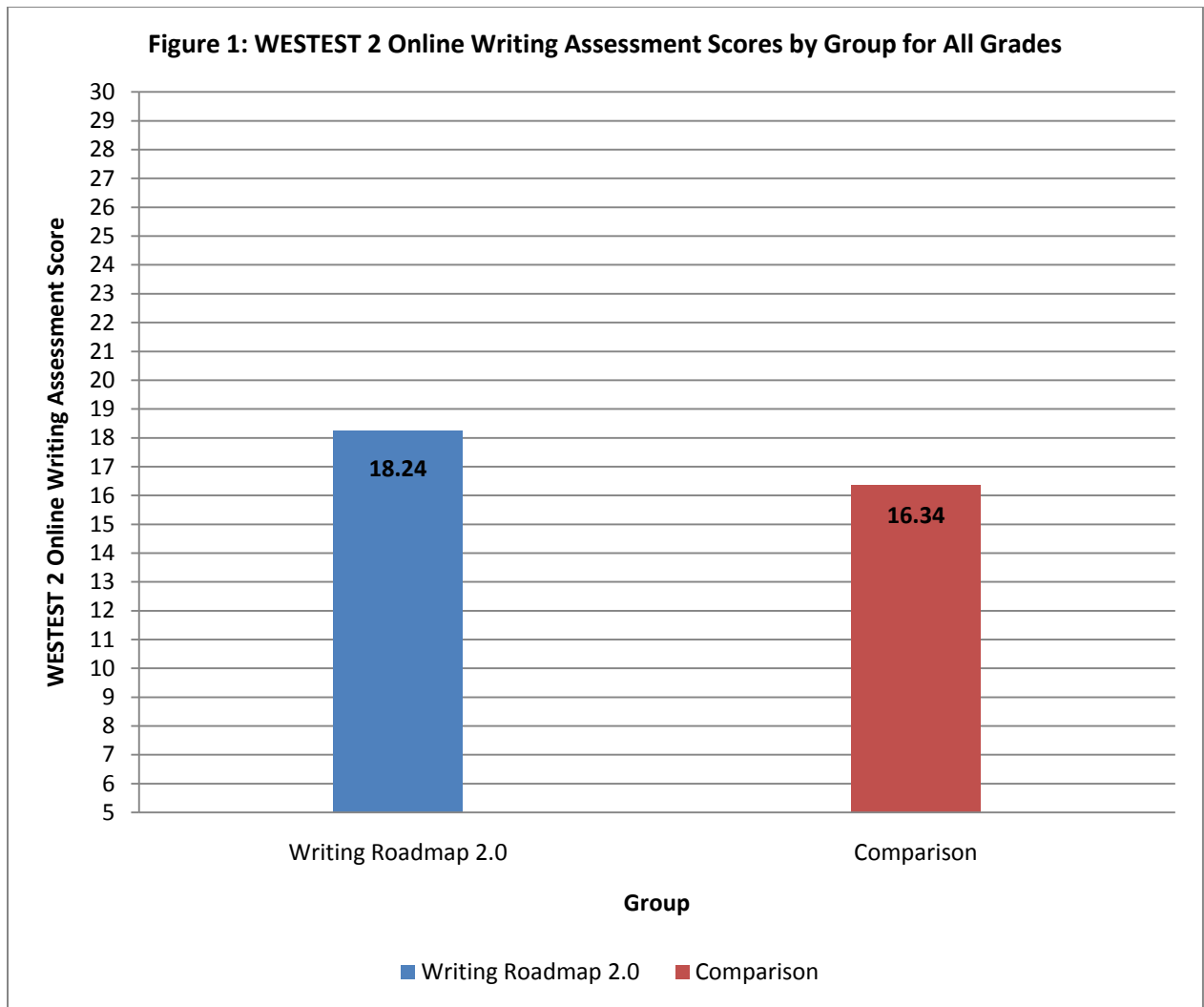
Initial analyses were completed with the entire sample of students to determine the overall impact of Writing Roadmap 2.0, irrespective of grade level.

Average Number of Writing Roadmap 2.0 Assignments Completed (All Grades)

Descriptive statistics were computed to determine the extent to which students in the treatment group used Writing Roadmap 2.0 during the 2008-2009 academic year. Results indicated the average number of times a student in the treatment group completed a Writing Roadmap 2.0 exercise was 6.70 ($SD = 3.04$).

Average WESTEST 2 Online Writing Assessment Scores (All Grades)

The range for the WESTEST 2 Online Writing Assessment in the current study was 25 points (i.e., between 5 and 30 points). While it is possible for a student to score zero points on the WESTEST 2 Online Writing Assessment, there were no scores of zero in either the comparison or treatment conditions for this study. Descriptive analyses indicated, as is illustrated in Figure 1, the average WESTEST 2 Online Writing Assessment score for the treatment group was 18.24 ($SD = 4.78$) and the average for the matched comparison group was 16.34 ($SD = 5.33$). This amounts to an average difference of approximately 2 more points on a 25 point scale for those students who used Writing Roadmap 2.0.



Significance Testing (All Grades)

An independent samples t-test was conducted to determine whether the difference between treatment and comparison students was statistically significant. Additionally, Cohen’s *d* was calculated to estimate the effect size of the relationship. Results appear in Table 3 below.

Table 3: T-Test Results for All Grades Combined

<i>df</i>	<i>t</i>	<i>p value</i>	Significance	Cohen’s <i>d</i>	<i>r</i>
16,660	24.26	<.000	YES	.38	.18

The results of the t-test indicate that there is a statistically significant difference in the WESTEST 2 Online Writing Assessment scores of students who completed five or more Writing Roadmap 2.0

exercises and similar students who did not, across all sampled grade levels. The effect size for this relationship was .38. This constitutes a small effect size according to Cohen (1960)³.

FINDINGS BY GRADE LEVEL

Next, researchers conducted statistical analyses within each sampled grade level (i.e., 4 through 9 and 11).

Average Number of Writing Roadmap 2.0 Assignments Completed (By Grade)

Table 4 illustrates that the average number of Writing Roadmap 2.0 assignments completed within the treatment group did not vary substantially by grade level. Students completed roughly 6 to 7 assignments. These findings mirror those from the aggregated sample ($M=6.70$; $SD=3.04$).

Table 4: Average Number of Writing Roadmap 2.0 Assignments Completed by Grade

Grade	<i>n</i>	Avg. Number of WRM 2.0 Assignments Completed	Std. Dev.
4	1,183	6.91	2.40
5	1,553	6.81	2.37
6	1,386	6.95	2.27
7	1,176	6.52	1.96
8	1,456	6.60	2.36
9	1,009	6.69	2.42
11	666	6.13	1.45

Average WESTEST 2 Online Writing Assessment Scores (By Grade)

As is indicated in Tables 5 and 6, the average WESTEST 2 Online Writing Assessment score for the treatment group was consistently higher than that of the comparison group. Figure 2 provides a graphical illustration of these grade level differences

³ Cohen (1988), provided the following interpretations of *d*: (.20 = small effect, .50 moderate effect, .80 = large effect)

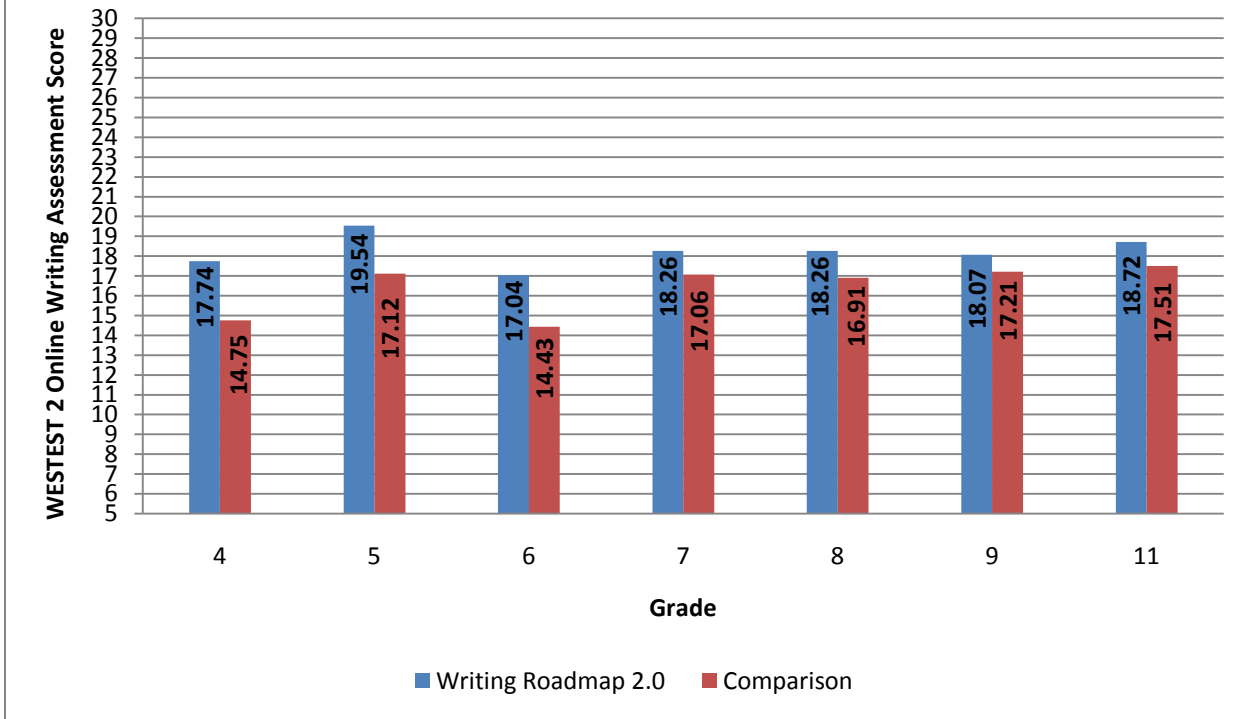
Table 5: Average WESTEST 2 Online Writing Assessment Scores for the Treatment Group by Grade

Grade	<i>n</i>	Avg. WESTEST 2 Online Writing Assessment Score	Std. Dev.
4	1,183	17.74	4.78
5	1,553	19.54	5.24
6	1,386	17.04	4.46
7	1,176	18.26	4.78
8	1,456	18.26	4.39
9	1,009	18.07	4.73
11	666	18.72	4.36

Table 6: Average WESTEST 2 Online Writing Assessment Scores for the Comparison Group by Grade

Grade	<i>n</i>	Avg. WESTEST 2 Online Writing Assessment Score	Std. Dev.
4	1,183	14.75	5.22
5	1,553	17.12	5.91
6	1,386	14.43	4.90
7	1,176	17.06	5.07
8	1,456	16.91	4.71
9	1,009	17.21	5.11
11	667	17.51	5.37

Figure 2: WESTEST 2 Online Writing Assessment Scores by Grade



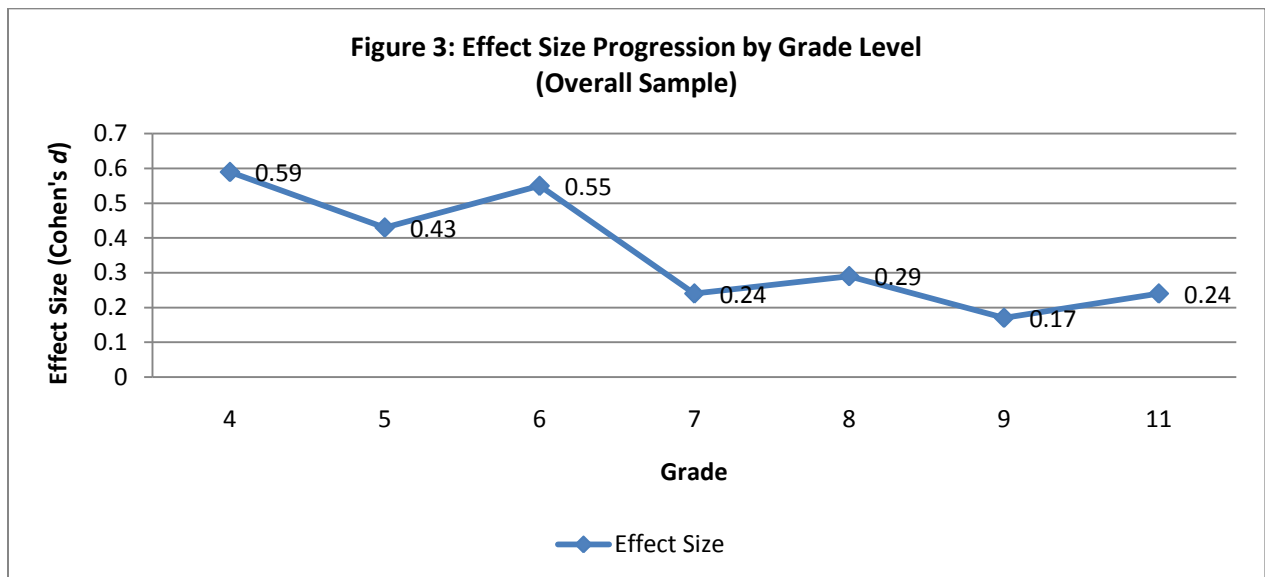
Significance Testing (By Grade)

A series of additional independent samples t-tests were conducted to determine whether the mean differences between groups were statistically significant for each sampled grade level. Additionally, Cohen's *d* was calculated to estimate the effect size of each relationship. Results appear in Table 7 below.

Table 7: T-Test Results by Grade

Grade	Mean Δ	<i>df</i>	<i>t</i>	<i>p value</i>	Significance	Cohen's <i>d</i>	<i>r</i>
4	+2.98	2,345	14.50	<.001	YES	.59	.28
5	+2.41	3,059	12.03	<.001	YES	.43	.21
6	+2.60	2,746	14.62	<.001	YES	.55	.26
7	+1.19	2,342	5.88	<.001	YES	.24	.12
8	+1.35	2,895	8.02	<.001	YES	.29	.14
9	+.86	2,004	3.92	<.001	YES	.17	.08
11	+1.21	1278	4.53	<.001	YES	.24	.12

All tests were statistically significant, indicating that for all sampled grades, students who completed five or more Writing Roadmap 2.0 assignments scored significantly higher on the WESTEST 2 Online Writing Assessment than a comparable group of students who did not use Writing Roadmap 2.0. The effect sizes for each relationship ranged between small and moderate, with the strongest effects present in grades 4, 5 and 6. This provides some evidence that Writing Roadmap 2.0, while effective for all students, has a greater impact on younger writers. The effect size trend by grade level is depicted in Figure 3 below.



REPLICATION WITH SMALLER RANDOM SAMPLES

Due to the large sample sizes utilized in this study, researchers decided to replicate significance testing within several smaller random samples to ensure the significance of these findings was not an artifact of sample size.

For the overall findings across sampled grade levels, five random samples containing 800 students each (i.e., 400 treatment and 400 control) were drawn from the larger sample and analyses were completed to determine the average effect size. Table 8 contains the significance testing, mean differences and effect sizes for each sample. A more detailed description can be found in Appendix A of this report.

Table 8: Significance and Average Effect Size across Five Random Samples for All Grades

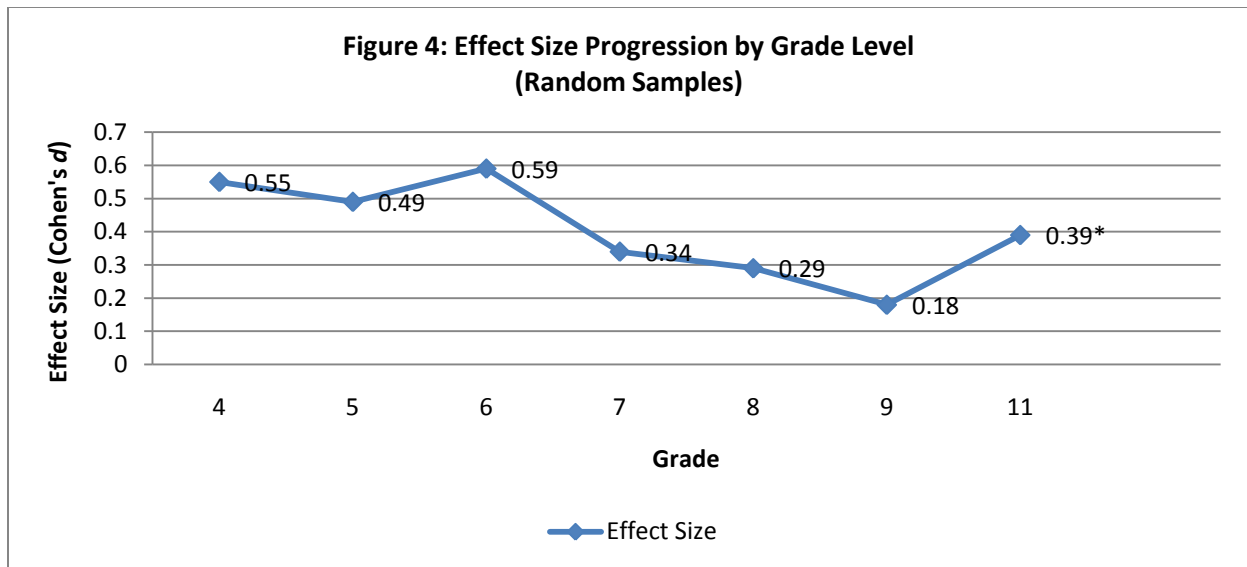
<i>Sample #</i>	<i>df</i>	<i>t</i>	<i>p value</i>	<i>Significance</i>	<i>Cohen's d</i>	<i>r</i>
1	798	4.51	.000	YES	.32	.16
2	787.43	3.67	.000	YES	.25	.13
3	798	4.69	.000	YES	.32	.16
4	798	3.31	.000	YES	.23	.12
5	798	3.31	.000	YES	.23	.12
AVG Effect Size					0.27	0.14

For the within grade level findings, seven random samples (i.e., one from each sampled grade) containing 800 students each (i.e., 400 treatment and 400 control) were drawn from the larger sample and analyses were completed to determine the effect size within each grade level. Table 9 contains the significance testing, and effect sizes found in each sample. Figure 4 presents the effect size progression by grade level. A more detailed description can be found in Appendix B of this report.

Table X: Significance and Effect Size for Seven Random Samples by Grade

<i>Sample #</i>	<i>df</i>	<i>t</i>	<i>p value</i>	<i>Significance</i>	<i>Cohen's d</i>	<i>r</i>
Grade 4	793.75	7.71	.000	YES	.55	.26
Grade 5	784.01	6.86	.000	YES	.49	.24
Grade 6	786.78	8.29	.000	YES	.59	.28
Grade 7	798	4.80	.000	YES	.34	.17
Grade 8	798	4.15	.000	YES	.29	.14
Grade 9	798	2.48	.000	YES	.18	.09
Grade 11 ⁴	754.37	5.44	.000	YES	.39	.19

⁴ Researchers caution against making generalizations to the larger sample using Grade 11 due to the possibility that the treatment and comparison groups are not closely matched on SES (See Appendix B)



Note: * See Footnote #3 on page 8.

CONCLUSION

The results of this initial study are quite promising for West Virginia. Data illustrate that there is a positive association between providing students with multiple opportunities to practice using Writing Roadmap 2.0 and their subsequent achievement on the WESTEST 2 Online Writing Assessment. These results were found in a large and geographically representative sample of WV students and replicated in smaller random samples. The magnitude of this relationship was strongest in the earlier grades and declined in the middle and high school grades. It is appropriate to note that this study used intact grade level cohorts and did not track the same students over time. This design is cross sectional in nature and provides a snapshot of student achievement for each group. However, a longitudinal design would increase internal control and likely yield more reliable data about the isolated impact of Writing Roadmap 2.0 on student achievement. Regardless, the results of the current study provide compelling evidence that, while the Writing Roadmap 2.0 program is effective for all students, it may be most beneficial when used with younger learners. It is important to note that a controlled experiment would be necessary to determine whether the differences in WESTEST 2 Online Writing Assessment scores found in this study are directly attributable to Writing Roadmap 2.0.

RECOMMENDATIONS

Given the positive association found in this preliminary study, it is recommended that Writing Roadmap 2.0 continue to be utilized as an instructional strategy for West Virginia students. Teachers of elementary grade children may find particular benefit from providing additional practice opportunities for these students. Furthermore, additional longitudinal research is necessary in order to fully isolate and evaluate the impact of Writing Roadmap 2.0 on student achievement. The planned expansion of the Writing Roadmap 2.0 program may create issues as researchers attempt to locate and track a suitable comparison group.

REFERENCES

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

APPENDIX A: REPLICATION ACROSS GRADE LEVELS

Five random samples containing 800 students each (i.e., 400 treatment and 400 comparison) were drawn from the larger Writing Roadmap 2.0 Study sample. After sampling, a series of ad-hoc descriptive analyses were conducted to ensure groups matched on socioeconomic status and geographic location (RESA membership was used as a proxy measure). Table A1 and figures A1 – A5 illustrate that the randomly selected treatment and comparison groups were closely matched on both variables.

Table A1: Socioeconomic Status Matching for the Five Random Samples

Group	Sample 1 % Low SES	Sample 2 % Low SES	Sample 3 % Low SES	Sample 4 % Low SES	Sample 5 % Low SES	AVG % Low SES
Treatment	48.5%	48.0%	47.3%	49.3%	52.5%	49.1%
Control	50.5%	49.5%	53.0%	47.8%	47.3%	49.6%

Figure A1: RESA Matching for Sample 1

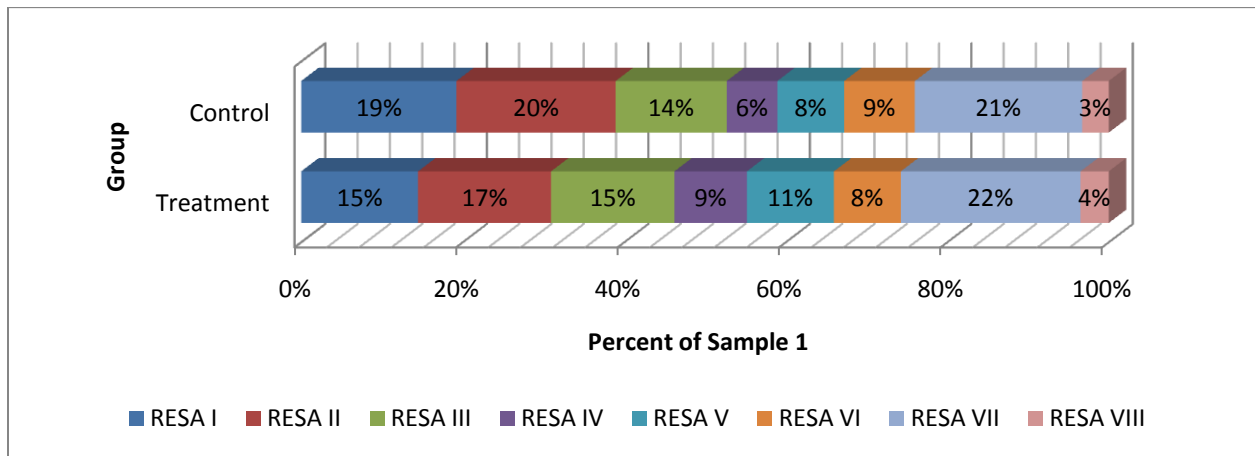


Figure A2: RESA Matching for Sample 2

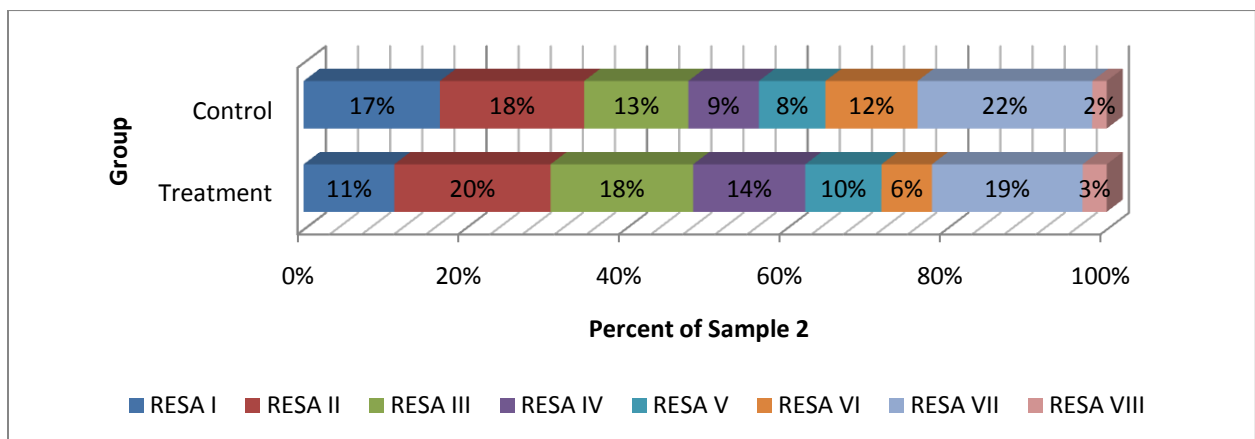


Figure A3: RESA Matching for Sample 3

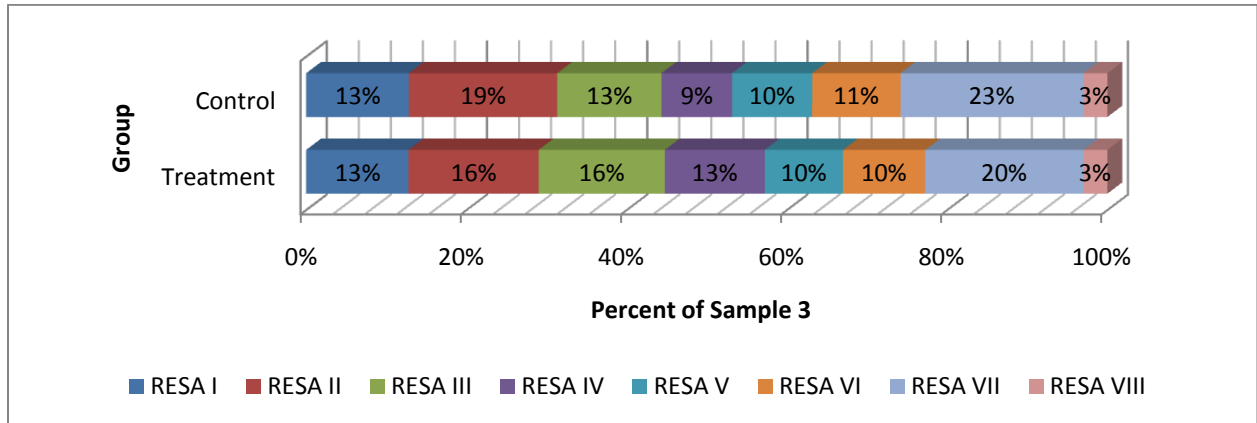


Figure A4: RESA Matching for Sample

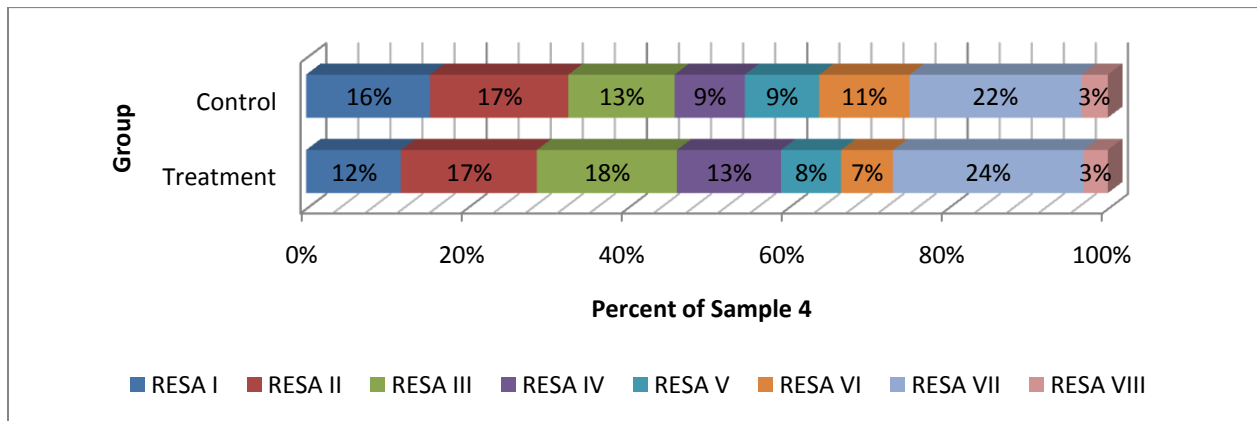


Figure A5: RESA Matching for Sample 5

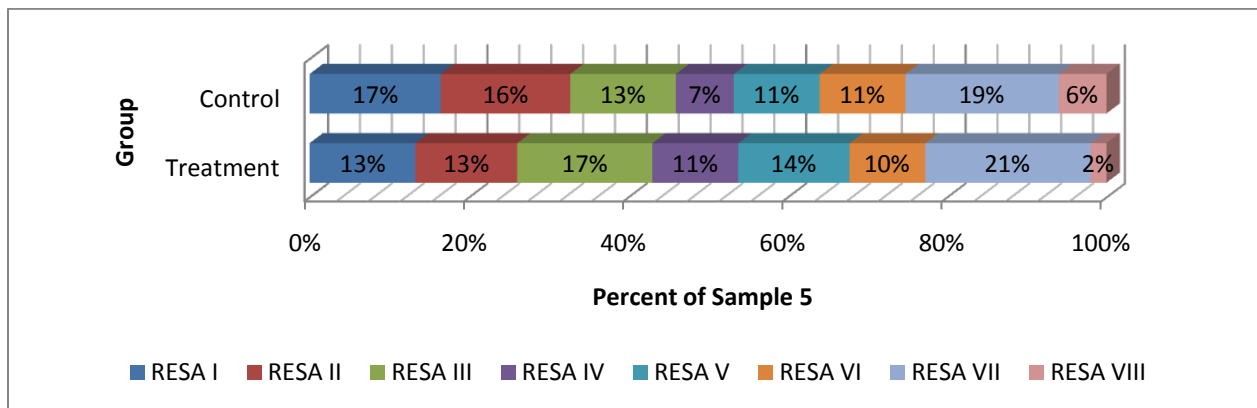


Table A2 illustrates the mean Composite Writing scores for both the treatment and control groups for each of the five random samples. In all cases, the treatment group scored marginally higher than the comparison group. The difference ranged from 1.22 points to 1.64 points.

Table A2: Mean Composite Writing Scores

Sample #	Treatment	Comparison	Mean Δ
1	17.99 SD =5.20	16.35 SD =5.10	1.64
2	18.22 SD =4.67	16.93 SD =5.24	1.29
3	18.20 SD =4.75	16.57 SD =5.13	1.63
4	18.04 SD =4.98	16.82 SD =5.41	1.22
5	18.04 SD =4.98	16.82 SD =5.41	1.22
AVG Mean Δ			1.40

A series of Independent Samples t-tests were conducted within each of the five random samples to determine whether mean differences in Table A2 were statistically significant. Table A3 illustrates that the differences were statistically significant for all samples ($p < .001$). As is noted at the bottom of Table A3, the average effect size for this relationship was .27 which constitutes a small positive effect for Writing Roadmap 2.0 according to Cohen (1988).

Table A3: T-Test Results

Sample #	df	t	p value	Significance	Cohen's d	r
1	798	4.51	.000	YES	.32	.16
2	787.43	3.67	.000	YES	.25	.13
3	798	4.69	.000	YES	.32	.16
4	798	3.31	.000	YES	.23	.12
5	798	3.31	.000	YES	.23	.12
AVG Effect Size					0.27	0.14

APPENDIX B: REPLICATION WITHIN GRADES

Seven random samples, one for each sampled grade (i.e., grades 4 - 9, and 11) containing 800 students each (i.e., 400 treatment and 400 control) were drawn from the larger Writing Roadmap 2.0 Study sample. After sampling, a series of ad-hoc descriptive analyses were conducted to ensure groups matched on socioeconomic status and geographic location. Table B1 illustrates that the randomly selected treatment and control groups were closely matched on socioeconomic status, with the exception of 11th grade where differences exceeded 20 percentage points. Figure B1 illustrates that the larger sample from which each grade level group was then drawn was geographically representative.

Table B1: Socioeconomic Status Matching for the Seven Random Samples by Grade

Grade	Grade 4 % Low SES	Grade 5 % Low SES	Grade 6 % Low SES	Grade 7 % Low SES	Grade 8 % Low SES	Grade 9 % Low SES	Grade 11 % Low SES	AVG % Low SES
Treatment	54.5%	51.4%	52.3%	50.9%	45.8%	47.9%	39.5%	48.9%
Control	45.5%	48.6%	47.8%	49.1%	54.3%	52.1%	60.5%	51.1%

Figure B1: RESA Matching for the Overall Sample

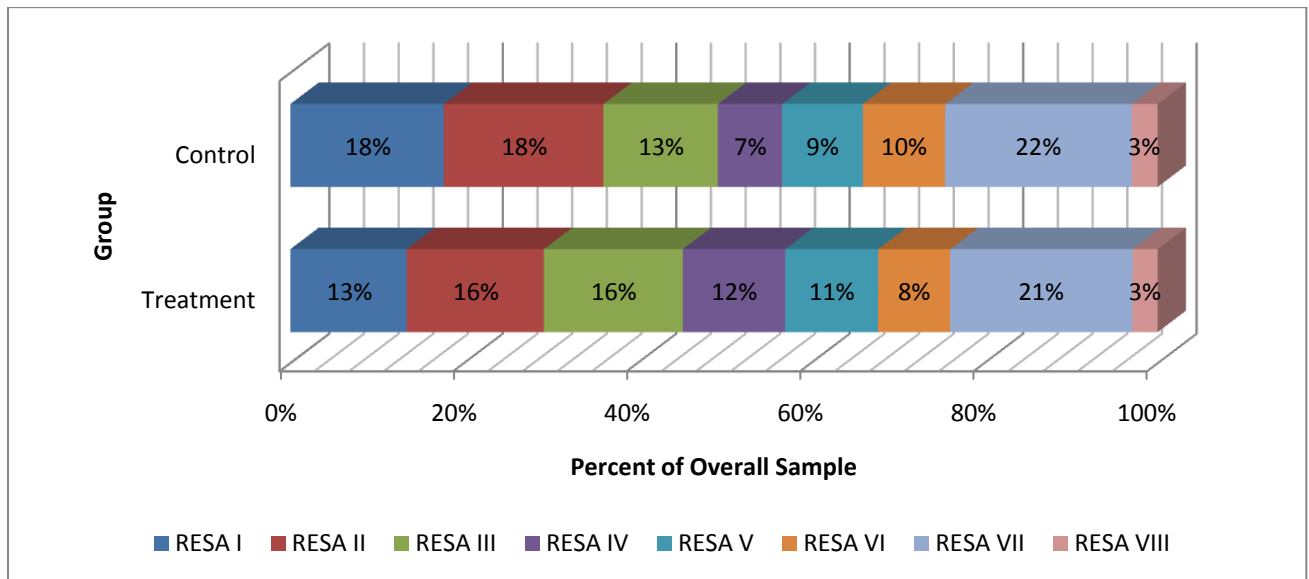


Table B2 illustrates the mean Composite Writing scores for both the treatment and comparison groups for each of the seven grade levels. In all cases, the treatment group scored marginally higher than the comparison group. The difference ranged from .88 points to 2.83 points.

Table B2: Mean Composite Writing Scores

Grade Level	Treatment	Comparison	Mean Δ
Grade 4	<i>M</i> = 17.57 <i>SD</i> = 4.85	<i>M</i> = 14.82 <i>SD</i> = 5.22	2.75
Grade 5	<i>M</i> = 19.35 <i>SD</i> = 5.26	<i>M</i> = 16.61 <i>SD</i> = 6.01	2.74
Grade 6	<i>M</i> = 17.09 <i>SD</i> = 4.51	<i>M</i> = 14.27 <i>SD</i> = 5.10	2.83
Grade 7	<i>M</i> = 18.30 <i>SD</i> = 4.87	<i>M</i> = 16.58 <i>SD</i> = 5.20	1.71
Grade 8	<i>M</i> = 18.40 <i>SD</i> = 4.45	<i>M</i> = 17.07 <i>SD</i> = 4.64	1.34
Grade 9	<i>M</i> = 17.96 <i>SD</i> = 4.85	<i>M</i> = 17.10 <i>SD</i> = 4.93	.88
Grade 11 ⁵	<i>M</i> = 19.14 <i>SD</i> = 4.30	<i>M</i> = 17.24 <i>SD</i> = 5.49	1.90
AVG Mean Δ			2.02

A series of Independent Samples t-tests were conducted within each of the five grade levels to determine whether mean differences in Table B2 were statistically significant. Table B3 illustrates that the differences were statistically significant for all grade levels ($p < .001$). Effect sizes ranged from very small (e.g., grade 9) to moderate (e.g., grades 4 and 6).

Table B3: T-Test Results

Grade Level	<i>df</i>	<i>t</i>	<i>p</i> value	Significance	Cohen's <i>d</i>	<i>r</i>
Grade 4	793.75	7.715	.000	YES	.55	.26
Grade 5	784.01	6.865	.000	YES	.49	.24
Grade 6	786.78	8.296	.000	YES	.59	.28
Grade 7	798	4.804	.000	YES	.34	.17
Grade 8	798	4.153	.000	YES	.29	.14
Grade 9	798	2.482	.000	YES	.18	.09
Grade 11 ¹	754.37	5.448	.000	YES	.39	.19

⁵ Researchers caution against making generalizations to the larger sample using Grade 11 due because the treatment and comparison groups are not closely matched on SES.