metis associates
making a meaningful difference

120 Wall Street
21st Floor
New York, NY 10005
Tel: 212-425-8833
Fax: 212-480-2176
www.metisassociates.com

Classroom, Inc.
2009-2013 Reading School Year Results in New York
Prepared by Metis Associates
May 2014

To assess student performance in reading, teachers implementing the Classroom, Inc. (CI) program during the past five years (2009 - 2013) administered a reading assessment, the *Reading-Level Indicator (RLI),* before and after the program was delivered (pretest/posttest design). This assessment was developed by the American Guidance Service, Inc. (now Pearson) and has been documented to be a valid and reliable measure of students' reading performance when used for the purpose of program evaluation.

For the past five years, Classroom, Inc. has contracted with Metis Associates to analyze and report on the reading performance of students participating in CI's school year programming. The analyses for this report combined the data from these five school years into a single dataset for students who took the *RLI* assessment.[1] The CI program is typically used to strengthen literacy skills during the school year, although the program also helps with improving student engagement and understanding the role of classroom education in preparation for career success. However, the focus of this study was to examine reading performance. The dataset includes records for 7,008 students who were assessed on the *RLI*.

Using this dataset, paired-samples *t*-tests were conducted on the scale scores to determine whether there were statistically significant differences in students' reading scores from pretest to posttest. Students who answered less than half of the items on either the pretest or the posttest were excluded from the analyses. Effect sizes were calculated using Cohen's *d* to determine the magnitude of the changes in reading performance. According to Cohen (1988), effect sizes of 0.2 are generally small, 0.5 are medium and 0.8 are large. In addition, analyses of covariance were conducted to determine whether there were differences in posttest scores across groups of students who completed 0 to 5 episodes, 6 to 9 episodes, or 10 or more episodes, after controlling for reading pretest scores. The results are provided in aggregate and also are disaggregated by implementation year, school affiliation, grade level, and gender.

[1] Two sites (N = 58 students) were removed from the analyses at the request of Classroom, Inc. for the purposes of creating a dataset that contained sites within New York City only.

RLI Results

From 2008 to 2013, a total of 7,149 students[2] completed both a pretest and posttest reading assessment. With the exclusion of the 141 students who had completed less than half of the items on either the pretest or posttest, the total number of students included in the analyses is 7,008.

Overall results were very positive and show that participating students experienced gains in their reading performance from an average pretest scale score of 114.98 to an average posttest scale score of 118.63. Gains were statistically significant at the .01 level and of an educationally meaningful and moderate magnitude (average effect size of 0.52). These results, along with the changes in reading performance for each year, are presented in Table 1. As shown in Table 1, during *each* school year from 2008-09 to 2012-13 students exhibited statistically significant gains in reading performance, with a medium effect size, indicating consistent moderate gains.

Table 1
School Years 2009-2013 Combined Analyses
***Reading-Level Indicator* Test Results**
Changes in Scale Scores, Across Years and by Implementation Year

Implementation Year	Matched N	Pretest Scale Score	Posttest Scale Score	Change in Scores (Posttest-Pretest)	*t* (Sig.) [a]	Effect Size (*Cohen's d*)[b]
2008-09	1,401	113.75	117.18	3.43	*18.945* (0.000)*	0.51
2009-10	1,725	114.83	118.69	3.86	*21.891* (0.000)*	0.53
2010-11	1,072	114.22	117.86	3.64	*17.929* (0.000)*	0.55
2011-12	1,850	115.83	119.69	3.86	*22.615* (0.000)*	0.53
2012-13	960	116.30	119.45	3.16	*14.453* (0.000)*	0.47
Total (2009 -13)	7,008	114.98	118.63	3.65	43.279 (0.000)*	0.52

[a] An asterisk in this column denotes a statistically significant difference at the p≤.01 level based on a paired-samples *t*-test.
[b] Effect size is a measure of the magnitude of the gains or losses, expressed in gain score standard deviation units. According to Cohen (1988), effect sizes of 0.2 are considered small, 0.5 are considered medium, and 0.8 are considered large. Effect sizes are presented only for statistically significant differences.

To make these results more readily understandable to the reader, the average scale scores at pretest and posttest were converted into grade equivalents (GE) of students' instructional reading level.[3] According to the test publisher, a reading scale score at pretest of 114.98 would correspond to an instructional reading level of 6.3, or sixth grade, third month, while a reading scale score at posttest of 118.63 would correspond to an instructional reading level of 7.8, or seventh grade, eighth month. Table A1 in the appendix presents the results of the grade-equivalent analyses for each implementation year.

To further assess changes over time, paired-samples *t* tests were conducted by school affiliation. Results show statistically significant gains for all three school affiliation groups: the Roman Catholic Diocese of Brooklyn, the Roman Catholic Archdiocese of New York, and the New York City Department of

[2] Note that no longitudinal analyses were conducted and students who participated in more than one year are treated as separate individuals for these analyses.

[3] As described in the publisher's manual, "grade equivalents are referred to as developmental norms because they place an individual along a span or continuum of development. Grade-equivalent values are presented in tenths of a grade." To calculate GEs, average scale scores were first converted to raw scores and then into a GE following the publisher's conversion tables provided in the manual. The average scale score at pretest (114.98) corresponds to a raw score of 25; the average scale score at posttest (118.63) corresponds to a raw score of 27. Note that GEs have many limitations. Since they are not equal-interval scales of measurement, they cannot be manipulated arithmetically (e.g., averaged) or used for direct longitudinal comparisons.

Education. The effect sizes by school affiliation ranged from 0.49 to 0.60, demonstrating moderate and educationally meaningful gains across the three groups. These results are presented in Table 2. Table A2 in the appendix presents the grade-level equivalent analyses for these disaggregated groups.

Table 2
School Years 2009-2013 Combined Analyses
Reading-Level Indicator Test Results
Changes in Scale Scores by School Affiliation

School Affiliation	Matched N	Pretest Scale Score	Posttest Scale Score	Change in Scores (Posttest-Pretest)	*t* (Sig.) [a]	Effect Size (*Cohen's d*)[b]
Diocese of Brooklyn	1,031	115.48	119.28	3.80	*18.026* (0.000)*	0.56
Archdiocese of NY	1,178	119.22	123.69	4.48	*20.701* (0.000)*	0.60
NYC DOE	4,799	113.84	117.25	3.41	*33.720* (0.000)*	0.49
All schools	7,008	114.98	118.63	3.65	*43.279* (0.000)*	0.52

[a] An asterisk in this column denotes a statistically significant difference at the p≤.01 level based on a paired-samples t-test.
[b] Effect size is a measure of the magnitude of the gains or losses, expressed in gain score standard deviation units. According to Cohen (1988), effect sizes of 0.2 are considered small, 0.5 are considered medium, and 0.8 are considered large. Effect sizes are presented only for statistically significant differences.

The data also were disaggregated by grade level and gender. As shown in Table 3, a statistically significant gain was observed among students in grades 4 through 9, who accounted for 98 percent of all students.[4] Most gains were moderate, with effect sizes between .50 and .74. .The most dramatic growth was for 4th grade students who were largely New York City public school students. The least dramatic growth, in grade 8, was still statistically significant. When looking at results by gender, the results also show that male and female students each experienced statistically significant gains, with female students demonstrating a gain slightly larger than male students.

Table 3
School Years 2009-2013 Combined Analyses
Reading-Level Indicator Test Results
Changes in Scale Scores by Grade Level and Gender

Student Characteristics		Matched N	Pretest Scale Score	Posttest Scale Score	Change in Scores (Posttest-Pretest)	*t* (Sig.)[a]	Effect Size (Cohen's d) [b]
Grade	Four	332	109.90	116.12	6.23	*13.454* (0.000)*	0.74
	Five	1,996	111.23	114.94	3.71	*24.999* (0.000)*	0.56
	Six	1,502	114.19	118.24	4.05	*24.227* (0.000)*	0.63
	Seven	1,212	117.48	121.05	3.56	*17.358* (0.000)*	0.50
	Eight	1,087	118.89	121.61	2.72	*11.660* (0.000)*	0.35
	Nine	767	118.87	122.18	3.31	*13.814* (0.000)*	0.50
Gender	Male	3,410	115.18	118.52	3.34	*26.412* (0.000)*	0.45
	Female	3,582	114.84	118.76	3.93	*35.074* (0.000)*	0.59

[a] An asterisk in this column denotes a statistically significant difference at the p≤.01 level based on a paired-samples t-test.
[b] Effect size is a measure of the magnitude of the gains or losses, expressed in gain score standard deviation units. According to Cohen (1988), effect sizes of 0.2 are considered small, 0.5 are considered medium, and 0.8 are considered large. Effect sizes are presented only for statistically significant differences.

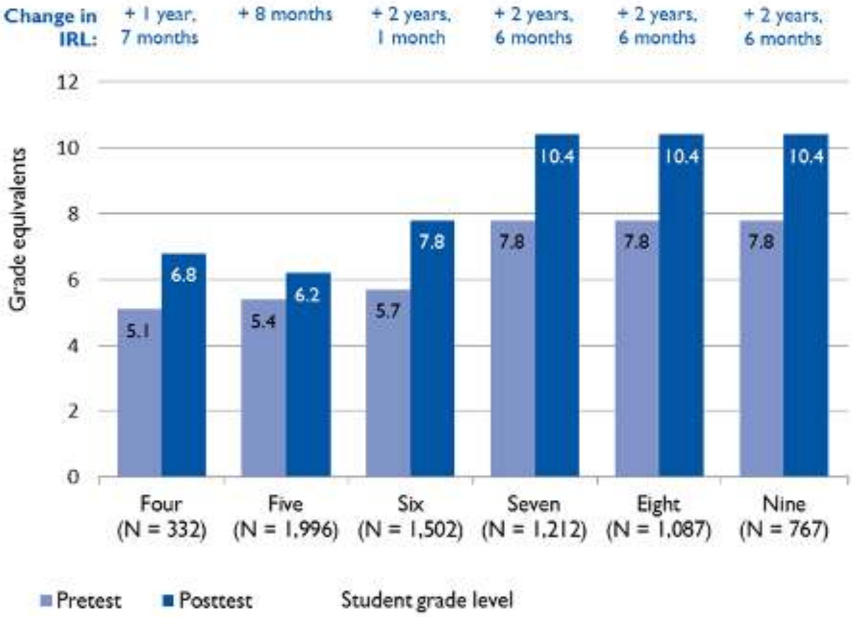
[4] The number of students in grades 10 (n=47), 11 (n=46) and 12 (n=13), included in the overall analyses, was too small to analyze separately, especially since most were from one high school. These students are omitted from Table 3, Figure 1, and in the Appendix, Tables A3 and A4.

The vast majority of the 4th and 9th grade test data come from NYC public schools. The test data for grades 5, 6, 7 and 8 is 62% from NYC public schools and 38% from students in schools of the Roman Catholic Archdiocese of New York and the Roman Catholic Diocese of Brooklyn.

To add to the understanding of the grade-level results, the average scale scores at pretest and posttest were converted into grade equivalents (GE) of students' instructional reading level. These grade level equivalents are presented in Figure 1. As this figure depicts, for students in grades 4, 6, 7, 8, and 9 the average increase in grade level equivalency over a school year was close to two years, with students in grades seven through nine demonstrating, on average, an increase of two years and six months. When examining the data in Figure 1, to better understand why the increase was lower for students in 5th grade, we found that 31% of the data in 5th grade comes from a single NYC public school. This school has traditionally started at a lower scale score (and hence, grade level) than most schools using our program, and has made the smallest gains.

These results, along with those for gender, are also presented in Table A3 in the appendix.

Figure 1
School Years 2009-2013 Combined Analyses
***Reading-Level Indicator* Test Results**
Changes in Grade Level Equivalencies by Grade Level



In addition, analyses of covariance were conducted to determine whether there were differences in posttest scores, after controlling for pretest scores, between groups of students who completed 0 to 5 episodes, 6 to 9 episodes, or 10 or more episodes.[5] As seen in Table 4, differences in adjusted posttest scores by number of episodes were statistically significant at the .01 level, although the magnitude of the difference was small. Post-hoc comparisons indicate that the only statistically significant difference is between those completing 0 to 5 episodes and those completing 10 or more episodes, thus pointing to the importance of completing most if not all of the episodes in a simulation.

[5] Viable data on the number of episodes completed was available for 6,386 students (91.1% of the 7,008 students with complete matched RLI test data).

Table 4
School Years 2009-2013 Combined Analyses
Reading-Level Indicator Test Results by Number of Episodes

Number of Episodes	Total N	Posttest Adjusted Mean Score[a]	F (Sig.)[b]	Effect size[c]	Post Hoc Comparisons
0 to 5	1,280	117.86	11.365 (0.000)*	0.13	[0 to 5]<[10 or more]
6 to 9	2,648	118.64			
10 or more	2,458	118.97			

[a] Posttest mean scores were adjusted to take into account pretest differences in W-ability scores.

[b] An asterisk denotes a statistically significant difference at the .01 level based on an analysis of covariance.

[c] Effect size is a measure of the magnitude of the gains or losses, expressed in gain score standard deviation units. According to Cohen (1988), effect sizes of 0.2 are considered small, 0.5 are considered medium, and 0.8 are considered large. Effect sizes are presented only for statistically significant differences.

Conclusions

The results from the combined school year 2009-2013 analyses indicate that there were statistically significant gains in the reading performance of students who, over the past five years participated in a Classroom, Inc. school year program. Results also indicate that students enrolled in schools from all three school affiliations—New York City Department of Education public schools, Roman Catholic Diocese of Brooklyn, and Roman Catholic Archdiocese of New York—improved their reading performance. Furthermore, the *RLI* findings indicate a relationship between the number of episodes completed and changes in reading performance, with those students completing 10 or more episodes demonstrating average gains larger than those completing five or fewer episodes during an academic year.

Appendix

Table A1
School Years 2009-2013 Combined Analyses
Reading-Level Indicator Test Results
Instructional Reading Level (IRL) by Implementation year

Implementation Year	Matched N	Pretest Scale Score	Posttest Scale Score	Pretest Instructional Reading Level	Posttest Instructional Reading Level	Change in Instructional Reading Level
2008-09	1,401	113.75	117.18	5.7	6.8	+ 1 year, 1 month
2009-10	1,725	114.83	118.69	6.3	7.8	+ 1 year, 5 months
2010-11	1,072	114.22	117.86	6.3	7.8	+ 1 year, 5 months
2011-12	1,850	115.83	119.69	6.9	9.5	+ 2 years, 6 months
2012-13	960	116.30	119.45	6.9	9.5	+ 2 years, 6 months
Total (2009-13)	7,008	114.98	118.63	6.3	7.8	+ 1 year, 5 months

Table A2
School Years 2009-2013 Combined Analyses
Reading-Level Indicator Test Results
Instructional Reading Level (IRL) by School Affiliation

School Affiliation	Matched N	Pretest Scale Score	Posttest Scale Score	Pretest Instructional Reading Level	Posttest Instructional Reading Level	Change in Instructional Reading Level
Diocese of Brooklyn	1,031	115.48	119.28	6.3	9.5	+ 3 years, 2 months
Archdiocese of New York	1,178	119.22	123.69	9.5	12.2	+ 2 years, 7 months
NYC DOE	4,799	113.84	117.25	5.7	6.8	+ 1 year, 1 month

Table A3
School Years 2009-2013 Combined Analyses
Reading-Level Indicator Test Results
Instructional Reading Level (IRL) by Student Characteristics

Student Characteristics		Matched N	Pretest Scale Score	Posttest Scale Score	Pretest Instructional Reading Level	Posttest Instructional Reading Level	Change in Instructional Reading Level
Grade	Four	332	109.90	116.12	5.1	6.8	+ 1 year, 7 months
	Five	1,996	111.23	114.94	5.4	6.2	+ 8 months
	Six	1,502	114.19	118.24	5.7	7.8	+ 2 years, 1 month
	Seven	1,212	117.48	121.05	7.8	10.4	+ 2 years, 6 months
	Eight	1,087	118.89	121.61	7.8	10.4	+ 2 years, 6 months
	Nine	767	118.87	122.18	7.8	10.4	+ 2 years, 6 months
Gender	Male	3,410	115.18	118.52	6.3	7.8	+1 year, 5 months
	Female	3,582	114.84	118.76	6.3	7.8	+1 year, 5 months