

**Pilot Study of
Millmark Education's
ConceptLinks[™] Ecosystems
Instructional Module**

July 3, 2007

Study Conducted and Report Submitted by:



Dr. Mary McNabb, Director
Learning Gauge Inc.

Dr. Charlene Cobb
Educational Consultant

Table of Contents

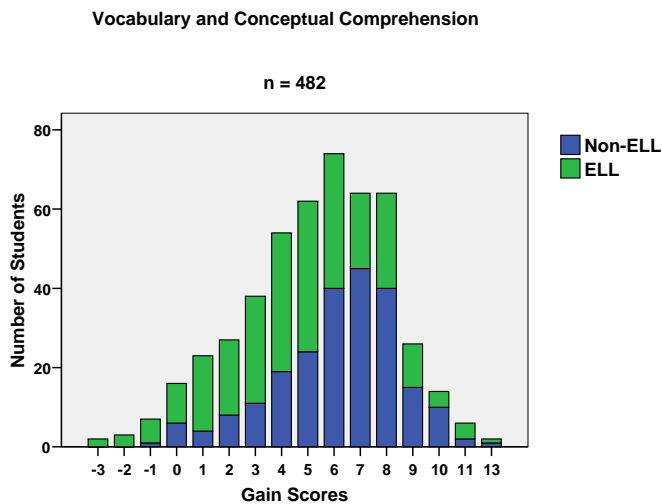
Executive Summary	3
Overview of the Pilot Study	4
Methodology	4
Statistical Results	7
Classroom Adoption Practices	13
Conclusions and Recommendations.....	16
Appendices	
Appendix A: Validation of the criterion-referenced test	19
Appendix B: Descriptive statistics.....	29

Executive Summary

A pilot study of the *ConceptLinks™ Ecosystems Instructional Module* was conducted by Learning Gauge, Inc. in partnership with Millmark Education during spring 2007. The purpose of the pilot study was to field test a prototype of Millmark Education's new instructional module with regard to its impact on students' content literacy and academic vocabulary, and teacher's ability to implement it effectively in the classroom. The study surveyed teachers about their instruction and the subsequent impact of Millmark's *ConceptLinks™ Ecosystems module* on students' content area knowledge and content literacy development. The module provides instructional resources for vocabulary building, concept development, and comprehension development along with language development. Comprehension development in the *ConceptLinks™ Ecosystems module* focused on determining importance when reading non-fiction texts. The materials are designed for multi-need instruction of students with Reading Levels 4 through 8 including grade level readers, English language learners (ELL), and struggling readers below grade level.

The pilot study involved 482 students and 11 teachers from nine public schools in urban and suburban settings in the Southeastern, Northeastern, Midwestern, Southwestern and Southern regions of the United States. Teachers were recruited to participate in the pilot study by Millmark Education and received 60 to 90 minutes of professional development from Millmark consultants. Duration of teacher training depended upon the size of the group. Students completed an average of 8 lessons with an average of 58 instructional minutes per lesson. The range of student gains was larger among struggling readers than among regular education students or those reading at grade level. There was not a direct correlation between type of student (Non-ELL or ELL) and reading at grade level. Of the 382 students reading at grade level, 44% of them were ELL students.

Chart 1. Learning gains in vocabulary and conceptual comprehension per student type.



Overall, students had a 33% gain in their vocabulary and conceptual comprehension. The mean difference in students' pre- and post-test performance on the combined vocabulary and conceptual comprehension subscales of the criterion-referenced test was 5.33, which is a statistically significant gain in student performance attributable to the instruction they received with the *ConceptLinks™ Ecosystems module*.

Overview of the Pilot Study

A pilot study of the *ConceptLinks™ Ecosystems* instructional module was conducted by Learning Gauge, Inc. in partnership with Millmark Education during spring 2007. The purpose of the pilot study was to field test a prototype of Millmark Education's new instructional module with regard to its impact on students' content literacy and academic vocabulary, and teacher's ability to implement it effectively in the classroom. The study surveyed teachers about their instruction and the subsequent impact of Millmark's *ConceptLinks™ Ecosystems* module on students' content area knowledge and content literacy development. The module provides instructional resources for vocabulary building, concept development, and comprehension development along with language development. Comprehension development in the *ConceptLinks™ Ecosystems* module focused on determining importance when reading non-fiction texts. The materials are designed for multi-need instruction of students with Reading Levels 4 through 8 including grade level readers, English language learners (ELL), and struggling readers below grade level.

Methodology

The pilot study for the module involved a two-fold design. First, Learning Gauge conducted a validation study of the criterion-referenced test developed by the Millmark Education consultant to measure students' vocabulary development, and comprehension of text features, visuals and science concepts. Students' pre-test results were used to conduct the validation study of the test consisting of 27 questions. Two questions were removed from the test due to low reliability of the items. The statistical results in this report are based on the remaining 25 test questions in four subscales. The four subscales are science vocabulary questions (7 items), non-fiction text feature questions (6 items), visual literacy or technical picture questions (3 items), and conceptual comprehension questions (9 items). For details about the validity and reliability of the measurement used in this study, refer to Appendix A.

Second, the pre-and post-test results from the validated measurement were statically analyzed for evidence of students' learning. Additional data about the instruction students received were gathered with a pre- and post-survey of participating teachers. Teachers were recruited to participate in the study through Millmark Education marketing consultants. Prior to initiating instruction within the *ConceptLinks™ Ecosystems* module, teachers attended one training session. This session varied from 60-90 minutes depending on the size of the group. Teachers also received all instructional materials as well as the pre-test and post-test at this time. Following the training session, teachers were asked to administer and score the pretest, then return all pretest materials for statistical analysis. After administration of the pretest, teachers initiated instruction. Teachers administered the post-test at the end of the instructional period, which ranged from 2 to 3 weeks in duration.

Characteristics of Treatment Groups

The pilot study involved 482 students and 11 teachers from nine public schools in urban and suburban settings in the Southeastern, Northeastern, Midwestern, Southwestern and Southern regions of the United States. Several teachers taught more than one class group. There were a total of 24 class groups. Of the 24 classes, 11 were science classes comprised of ELL only and 13 were science classes consisting of a mix of non-ELL and ELL. Three of the classes were small multi-grade level groups and 21 were grade level groups. This pilot study did not have a control group, which was a decision made by Millmark Education. Teachers were recruited to participate in the pilot study by Millmark Education consultants. Two teachers who piloted the instructional module and completed the pre- and post-survey conducted individual lessons with students who were pulled out of their classes for remedial purposes. Those students' pre- and post-test data is not included in these statistical analyses, although their teacher survey data is included in the Qualitative section of this report as feedback to Millmark Education.

Students in the treatment class groups completed an average of 8 lessons with an average of 58 instructional minutes per lesson, although individual teachers varied in how they implemented the instructional module. Table 1 shows the number of lessons and the minutes of instruction each teacher provided students for each lesson.

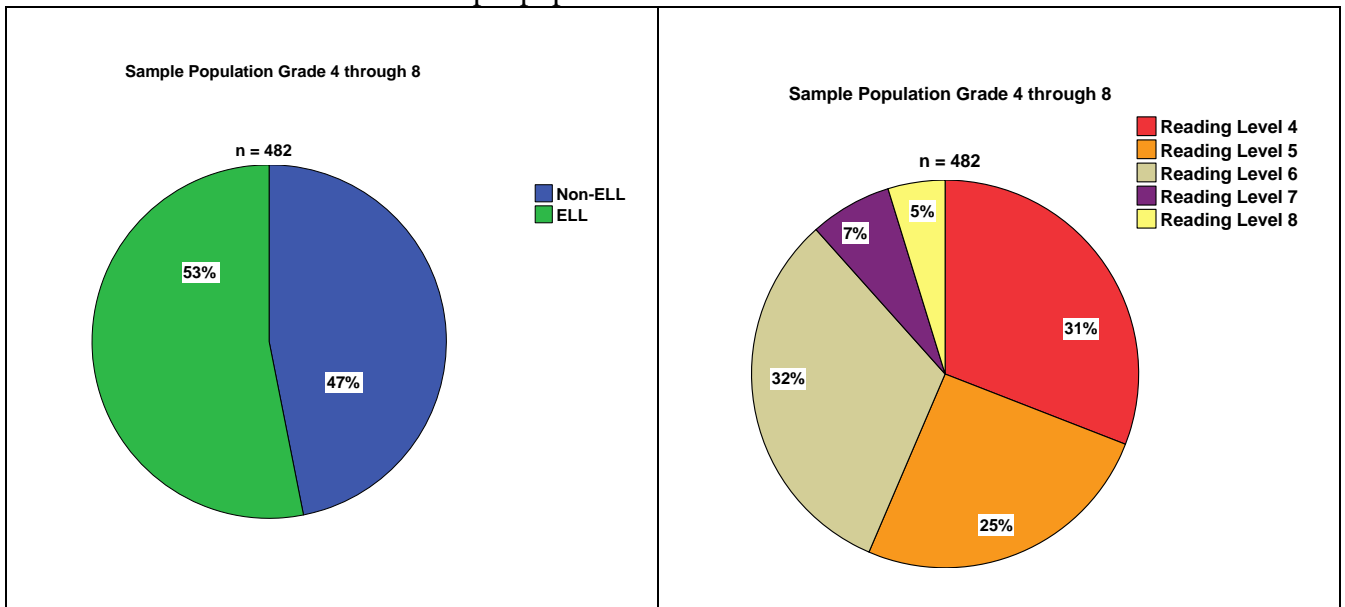
Table 1. Weekly instructional times and lessons completed for each teacher.

Teacher	Sessions per week	Min per session	Minutes per week	Lessons completed
1	4	45	180	10
2	5	60	300	6
3	5	45	225	10
4	2	30	60	10
5	4	70	280	10
6	3	132	396	8
7	5	45	225	10
8	2	90	180	3
9	2	105	210	10
10	5	38	190	10
11	5	42	210	8
Totals	46		2596	100
Average number of minutes per session				56.43
Total minutes of instruction				5643.48
Average number of lessons completed				7.69
Average number of minutes per teacher				434.11
Average number of minutes per lesson				54.26

The One-sample Kolmogorov-Smirnov (1-Sample K-S) test, Kaiser-Meyer-Olkin Measure of Sampling Adequacy and the Bartlett's test of sphericity tests were conducted on Millmark Education's sample of students. These data showed that all participant groups fall within the normal population distribution and meet the requirements for comparability of the groups. A few outliers were found with the data and these student cases were deleted from the data set analyzed for this report. Student cases were also deleted when there was no matched pair pre- and post-test due to student mobility. The sample population, although normal, did show a wide range and variability on pre-test performance.

Of the 482 students with matched pair pre- and post-tests, 47% were non-ELL students and 53% were identified as ELL students. The ELL students were not necessarily struggling readers. Twenty-one percent of all students were struggling readers, both non-ELL and ELL, who read between .5 and 2.5 levels before their grade level. Of the 226 native English learners (non-ELL) 96% read at grade level. Of the 256 English language learners 68% read at their grade level. The 482 students in the treatment groups ranged from grades 4 through 8. The pie charts in Chart 2 shows the distribution of the student sample based on type of learner (ELL or Non-ELL) and grade levels.

Chart 2. Characteristics of the sample population.



Teachers in the treatment classrooms were assigned to teach 10 lessons within a two week time span. There were a total of 10 lessons in the instructional module. Actual implementation among treatment teachers varied from two to three weeks due to block scheduling and other unknown differences. Total number of lessons that teachers completed also varied according to the post-survey data compiled in Table 1.

Statistical Results

Overall, results of the pilot study show that the 482 students had a mean gain of 2.20 across all test questions. The Paired Samples T-Test result for these students is $t(481) = 14.086$; $p < 0.01$, which indicates statistically significant gains in students' performance on the criterion-referenced test (see Table 2). The effect size is $d = .45$ ($r = 0.22$), which indicates the size of the effect is small. Although students' overall gain is statistically significant, 34% of those 482 students showed negative or zero gain on the post-test while 66% of students showed gains ranging from 1 to 11 points on the post-test (for details see Appendix B Frequency Tables). Due to the wide range in student performance on the pre- and post-tests, Two-Step Cluster Analyses (Bayesian Criterion) were conducted to identify whether or not there was a reading level effect on student groups' learning outcome.

Table 2. Paired Samples Test for all the test questions.

Paired Differences between Pre- and Post-Test Scores					t	df	Sig. (1-tailed)
Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
			Lower	Upper			
2.199	3.428	.156	1.892	2.506	14.086*	481	.0005

* indicates significant difference between pre- and post-test performance when $p < 0.01$

Chart 3 shows the results of the student clusters and means for each cluster. Students at reading levels 7 and 8 had the widest range and the lowest average gains from the instructional module. The instructional module is written within the range of grades 3-5, and so readability of actual text should not be a factor in student performance for those at reading level 7 and 8. As such, the reason for the low performance among 7th and 8th grade regular education students is unknown.

The other three reading level clusters are comprised of a mix of grade level readers and readers below their grade level. These student clusters also span both native English learners and English language learners. Students at Reading Level 4 (which correlates to Grade 4) showed the greatest gains according to the cluster analysis. The cluster analysis indicates that students at Reading Level 4 had a mean gain of 2.79 while students at Reading Levels 5-8 had a mean gain of 1.93.

Chart 3. A result from Bayesian cluster analysis indicates differences in learning among students at the various reading levels.

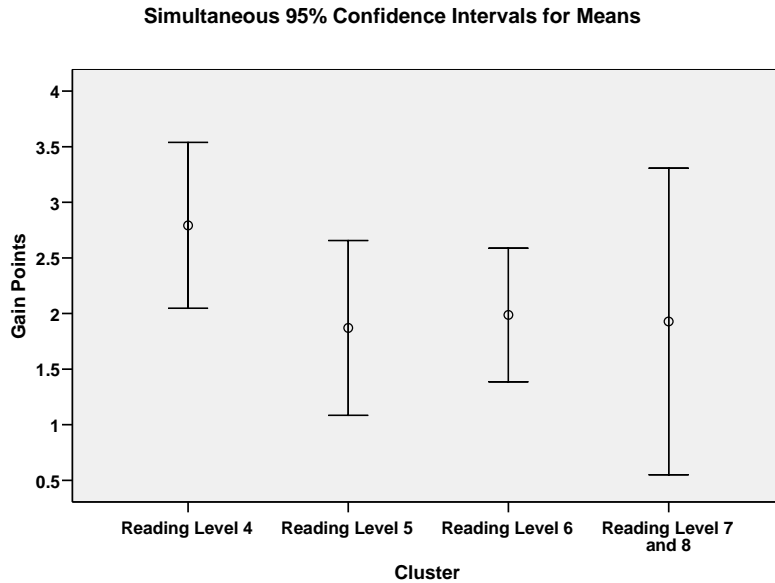
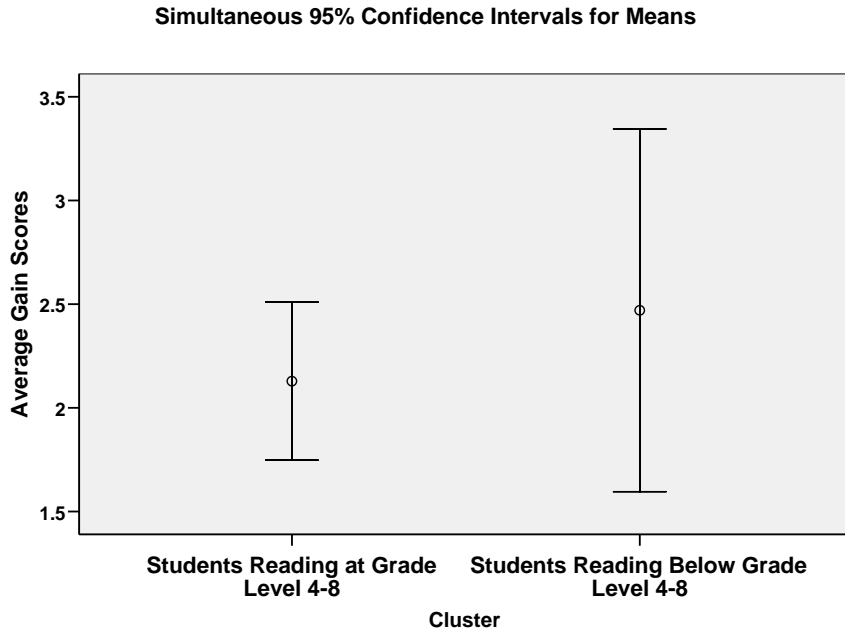


Chart 4 shows the results of a second cluster analysis. It indicates that the average gain for grade level readers was 2.13 and 2.47 for struggling readers across all grades. The range of student gains was larger among struggling readers (1.6 to 3.4 points on the test) than among regular education students or those reading at grade level (1.7 to 2.5 points on the test). Of participating students, 79% were reading at grade level, 2% were 0.5 below grade level, 2% were 1.5 below grade level and 17% were 2.5 below grade level in reading. There was not a direct correlation between type of student (Non-ELL or ELL) and reading at grade level. Of the 382 students reading at grade level, 168 (44%) of them were ELL students. The remaining 88 ELL students did comprise 88% of those reading below grade level, 86% of who were 2.5 below grade level. In this regard, the data suggests that the instructional module provides effective opportunities to learn for all students and is particularly helpful to struggling readers.

Chart 4. Cluster analysis indicates that the 21% of struggling readers in the study gained more than regular education students.



Exploratory statistics were conducted on the four subscale on the criterion-referenced test. Analysis of these data indicates that student gains clustered around two of the four subscales in the pre- and post-test. Students had the most learning gains for the vocabulary questions and conceptual comprehension questions (see Table 3). Charts 5 through 8 indicate the learning gains for each subscale of the criterion-referenced test.

Chart 5. Vocabulary subscale.

Chart 6. Conceptual comprehension

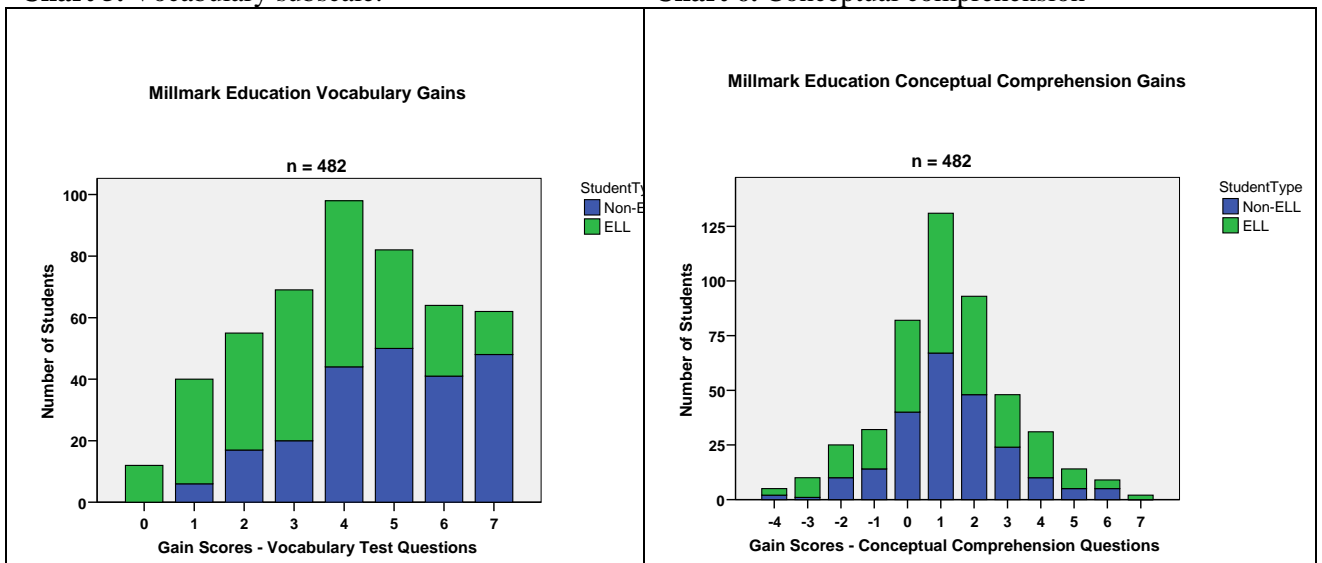


Chart 7. Gains for text feature subscale.

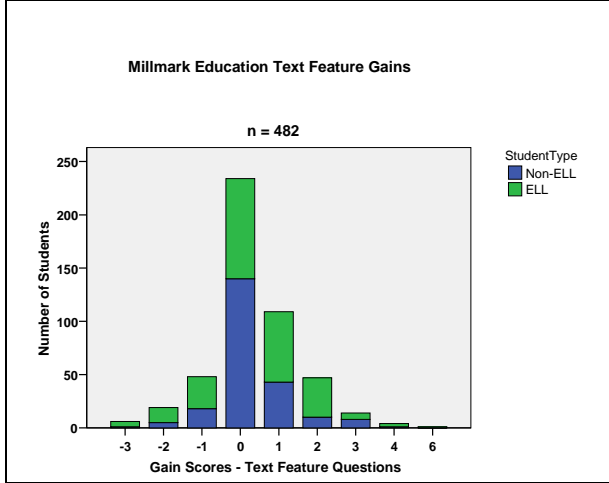


Chart 8. Gains for visual literacy subscale.

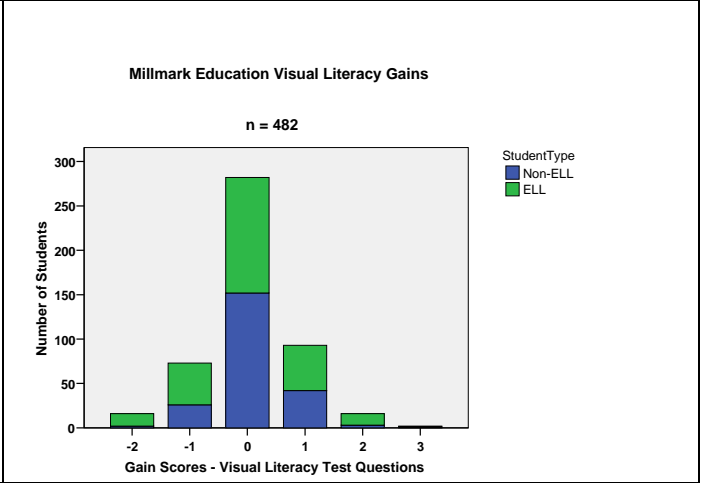


Table 3. Shows the statistical results that coincide the data in Charts 5 through 8.

	N	Mean	Std. Deviation
Concept Gain	482	1.23	1.924
Visual Gain	482	.05	.804
Text Feature Gain	482	.34	1.159
Vocabulary Gain	482	4.10	1.898

Some 99.5% of students gained in vocabulary, 68% gained in conceptual comprehension, 36% gained in text features, and 23% gained in visual literacy. Further analysis of students' learning gains at each grade level and for each teacher indicate that all class groupings showed a positive result for the combined vocabulary and conceptual comprehension subscales of the test. Aggregated class group results in Table 4 were all statistically significant gains. Data for students reading below grade level are indicated. Although 34% of participating students had a negative or zero gain on the post-test, these data do not show a strong pattern of teacher effect. The 34% of students whose pre- and post-test differences did not indicate a learning gain were in grades 4 through 8. Five teachers (2, 5, 6, 7, and 8) had the majority of students with no gains. However, there is no clear correlation between classroom use of the instructional module as reported by teachers on the post-survey and student gains.

Table 4. Shows the average gains in vocabulary and conceptual comprehension for each teacher group and reading level of students.

Grade 4							
Teacher	Reading	Mean	N	% of Total Sum	Variance	Std. Error of Mean	Std. Deviation
1	At Grade Level	6.48	21	19.3%	9.462	.671	3.076
2	At Grade Level	4.91	78	54.4%	8.317	.327	2.884
3	At Grade Level	5.97	29	24.6%	7.034	.493	2.652
	0.5 Below Grade Level	4.00	3	1.7%	3.000	1.000	1.732
Total for 4th Grade		5.37	131	100.0%	8.390	.253	2.897
Grade 5							
Teacher	Reading	Mean	N	% of Total Sum	Variance	Std. Error of Mean	Std. Deviation
4	At Grade Level	6.20	20	24.3%	4.484	.474	2.118
	0.5 Below Grade Level	4.00	2	1.6%	32.000	4.000	5.657
5	At Grade Level	5.84	64	73.2%	6.166	.310	2.483
	0.5 Below Grade Level	.25	4	.2%	3.583	.946	1.893
	1.5 Below Grade Level	4.00	1	.8%	.	.	.
Total for 5th Grade		5.62	91	100.0%	7.217	.282	2.686
Grade 6							
Teacher	Reading	Mean	N	% of Total Sum	Variance	Std. Error of Mean	Std. Deviation
6	At Grade Level	6.87	82	67.7%	2.932	.189	1.712
	0.5 Below Grade Level	4.00	2	1.0%	18.000	3.000	4.243
	1.5 Below Grade Level	5.71	7	4.8%	5.571	.892	2.360
	2.5 Below Grade Level	4.00	12	5.8%	7.636	.798	2.763
7	At Grade Level	3.29	17	6.7%	6.971	.640	2.640
9	At Grade Level	5.33	6	3.8%	5.467	.955	2.338
	2.5 Below Grade Level	3.00	5	1.8%	4.000	.894	2.000
10	At Grade Level	7.86	7	6.6%	7.810	1.056	2.795
11	At Grade Level	5.00	3	1.8%	.000	.000	.000
Total for 6th Grade		5.90	141	100.0%	6.361	.212	2.522
Grade 7							
Teacher	Reading	Mean	N	% of Total Sum	Variance	Std. Error of Mean	Std. Deviation
7	At Grade Level	3.05	21	28.7%	5.848	.528	2.418
8	2.5 Below Grade Level	3.88	26	45.3%	11.146	.655	3.339
9	At Grade Level	7.00	1	3.1%	.	.	.
	0.5 Below Grade Level	2.00	1	.9%	.	.	.
10	At Grade Level	3.75	4	6.7%	1.583	.629	1.258
11	At Grade Level	5.67	6	15.2%	1.867	.558	1.366
Total for 7th Grade		3.78	59	100.0%	7.864	.365	2.804

Table 4. (continued) Shows the average gains in vocabulary and conceptual comprehension for each teacher group and reading level of students.

Grade 8							
Teacher	Reading	Mean	N	% of Total Sum	Variance	Std. Error of Mean	Std. Deviation
7	At Grade Level	4.43	21	31.2%	4.057	.440	2.014
8	2.5 Below Grade Level	5.08	37	63.1%	8.132	.469	2.852
10	At Grade Level	6.00	1	2.0%	.	.	.
11	At Grade Level	11.00	1	3.7%	.	.	.
Total		4.97	60	100.0%	7.084	.344	2.662

A paired samples T-Test was conducted on the difference between students' pre- and post-test scores on the vocabulary and conceptual comprehension subscales. The paired sample T-Test result is $t(481) = 42.127$; $p < 0.01$, which is a statistically significant gain.

Table 5. Combined Vocabulary and Conceptual comprehension Paired Samples T-Test.

Combined Vocabulary and Conceptual Comprehension	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
	5.33	2.777	.126	5.079	5.576	42.127*	481	.0005

* indicates significant difference between pre- and post-test performance when $p < 0.01$

Students completed an average of 8 lessons with an average of 58 instructional minutes per lesson. Overall, students had a 33% gain in their vocabulary and conceptual comprehension. The mean difference in students' pre- and post-test performance was 5.33, which is a statistically significant gain in student performance attributable to the instruction they received with the *ConceptLinks™ Ecosystems* module. Students also showed a significant gain in text features and no significant gain in visual literacy.

Table 6. Text Features Paired Samples T-Test.

Non-Fiction Text Features Subscale	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
	.338	1.159	.053	.234	.442	6.407*	481	.0005

* indicates significant difference between pre- and post-test performance when $p < 0.01$

Table 7. Visual literacy Paired Samples T-Test.

Visual Literacy Subscale	Paired Differences					t	df	Sig. (1-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
	.054	.804	.037	-.018	.126	1.474	481	.070

Classroom Adoption Practices

Participating teachers were asked to complete both pre- and post-field test surveys. As part of the pre-survey, teachers reported their level of knowledge and practice on several areas related to effective instruction. The areas included academic vocabulary, reading strategies for improving comprehension, expository text comprehension, expository writing, differentiation of instruction, flexible grouping, and using assessment to guide instruction. Teacher level of knowledge and practice was reported according on a 4-point continuum of 4) expert, 3) skilled, 2) capable, or 1) inexperienced.

A total of 10 teachers completed the pre-survey. For nearly every area, with the exception of expository writing, teachers reported their level of knowledge and practice to be in the range of expert to skilled. For expository writing, teacher reported their level between skilled and capable. The average level reported by teachers in each of the areas was academic vocabulary 3.5, reading strategies for improving comprehension 3.2, expository text comprehension 3.2, expository writing 2.8, differentiation of instruction 3.4, flexible grouping 3.4, and using assessment to guide instruction 3.5. Table 8 below provides a profile of each teacher that completed the pre-survey.

Table 8. Teacher profiles from pre-survey data.

1	Expert in assessment to guide instruction; skilled in all other areas
2	Inexperienced in expository writing; capable in assessment to guide instruction; skilled in all other areas
3	No pre-survey completed by this teacher
4	Inexperienced in expository writing; capable in expository reading strategies; skilled in comprehension strategies, differentiation, and grouping for instruction; expert in academic vocabulary and assessment to guide instruction
5	No pre-survey completed by this teacher
6	Skilled in expository writing; expert in all other areas
7	Skilled in all areas
8	Expert in academic vocabulary, differentiation, and assessment to guide instruction; skilled in all other areas

9	Expert in grouping for instruction; capable in comprehension strategies; skilled in all other areas
10	Expert in all areas
11	Expert in all areas
12	No pre-survey completed by this teacher
13	Skilled in all areas

In the area of academic vocabulary 5 teachers (6, 4, 8, 10, 11) reported their level as expert, and 5 teachers (1, 2, 7, 9, 13) reported their level as skilled. In the area reading strategies for improving comprehension 3 teachers (6, 10, 11) reported their level as expert, 6 teachers (1, 2, 4, 7, 8, 13) reported their level as skilled, and 1 teacher reported (9) the level as capable. With regard to expository text comprehension 3 teachers (6, 10, 11) reported their level as expert, 6 teachers (1, 2, 7, 8, 9, 13) reported their level as skilled, and 1 teacher (4) reported the level as capable. In terms of expository writing 2 teachers (1, 11) reported their level as expert, 6 teachers (1, 6, 7, 8, 9, 13) reported their level as skilled, and 2 teachers (2, 4) reported their level as inexperienced. For differentiation of instruction 4 teachers (6, 8, 10, 11) reported their level as expert and 6 teachers (1, 2, 4, 7, 9, 13) reported their level as skilled. In terms of grouping students for instruction 4 teachers (6, 9, 10, 11) reported their level as expert and 6 teachers (1, 2, 4, 7, 8, 13) reported their level as skilled. Finally, in using assessments to guide instruction 6 teachers (1, 4, 6, 8, 10, 11) reported their level as expert; 3 teachers (7, 9, 13) reported their level as skilled; and 1 teacher (2) reported a level of capable.

Prior to initiating instruction within the field test, teachers attended one training session. This session varied from 90-120 minutes depending on the size of the group. Teachers also received all instructional materials as well as the pretest and posttest at this time. Following the training session, teachers were asked to administer and score the pretest, then return all pretest materials for analysis and evaluation. After administration of the pretest, teachers initiated instruction.

Following the field test, teachers were asked to complete a post survey. A total of 13 teachers provided responses to an array of questions related to implementation of the *ConceptLinks™ Ecosystems* module. More than half of the teachers (7) completed all ten lessons in the module. One teacher completed 9 and 2 completed 8 lessons. Two other teachers completed 6 lessons and one teacher completed just 3 lessons. Most of the teachers (8) used all four levels of books in the *ConceptLinks™ Ecosystems* module with small groups of students in their classrooms. Only 1 teacher used a single level of text, and the remaining teachers used either 2 or 3 levels of texts. Seven teachers also reported that students read at least one additional level of text during instruction with the module.

Overall, teachers rated the Millmark educational program as useful in helping to achieve their district instructional goals. Most of the teachers (11 of 13) felt that the expectation of what could be accomplished in one lesson was appropriate for the duration of one instructional session and that the overall instructional plan was meaningful for their

students. These teachers also reported that the small group instructional plan was effective.

Teachers were also asked to provide additional information, reporting their responses as 5) excellent, 4) good, 3) average, 2) fair, or 1) poor. Teachers reported on specific components of the *ConceptLinks™ Ecosystems* module, including the student books, the teacher's guide, and the Concept Connector. Teachers reported that the student books were excellent to good (4.3 out of 5.0). Teachers commented that the books were highly engaging, providing pictures and simple language for each level of proficiency. Teachers reported that the teacher's guide was good (4.0 out of 5.0). Some teachers felt that the guide was well organized, well-paced, and generally easy to follow. Other teachers stated that the guide was confusing at first with the placement of lessons 6-10 at the bottom of the pages for lessons 1-5.

The Concept Connector was reported as good to average (3.8 out of 5.0). Teachers reported that the students enjoyed the Concept Connector and found it very useful as a way to activate prior knowledge before reading. Suggestions for improving the Concept Connector included making it larger with a separate teacher flap, possibly poster size and having it be more "technology friendly" for use with a smart board.

Most of the teachers (11 out of 13) reported that the *ConceptLinks™ Ecosystems* module was effective for teaching reading, as well as teaching science. Teachers would use it again as designed or with some modifications. Teachers stated that because the books were interesting, students wanted to read more. They also stated that the activities were engaging and aligned to district standards of learning. The fact that the books were written in clear and precise language enabled the students to comprehend the content and respond well to the instruction. The books were also reported as an excellent resource for integrating science instruction with hands-on learning in a science lab.

With regard to assessment, teachers reported that the Millmark instructional program assessments were useful for informing their instruction as well as informing students about their learning progress. Teachers stated that the leveled assessments enabled students to evaluate their knowledge on information learned at their level. They found the assessments to be well-written, helpful in clarifying which concepts needed reteaching, and useful in gauging student understanding. Teachers also reported that the Millmark study guides and black line masters both useful and helpful. Several teachers stated that the students found them engaging and it helped the teachers to focus and guide instruction.

Teachers were asked to provide insights about using Millmark in three specific areas: comprehension strategy instruction; vocabulary and academic knowledge development; writing instruction and student writing products. With regard to comprehension strategy instruction several teachers mentioned the value of text features in the *ConceptLinks™ Ecosystems* student books, i.e. titles, heading, photographs, captions, and bold print as an aid to comprehension. They felt that the comprehension strategy instruction enabled students to assess their own learning and apply fix-up strategies.

All teachers found the vocabulary and academic knowledge development in the *ConceptLinks™ Ecosystems* module to be particularly well done. Many teachers mentioned that the vocabulary focus in all student books was excellent. Some features of the vocabulary instruction that they found most useful were focusing on key vocabulary before reading, using visuals to determine the meaning of the word, and on page definitions of key vocabulary.

Due to the time frame of the field test in various locations, some teachers were unable to complete all of the writing instruction and student writing products. Those teachers who were able to incorporate writing found that the several positive aspects including, the variety of writing activities, the focus on different purposes for writing, and the use of writing to reinforce vocabulary and academic concepts.

When asked about the strengths of the Millmark instructional program, nearly all teachers mentioned the leveled books as a primary strength. Teachers felt that use of a common theme with differentiated texts was good for all learners, but particularly effective with ELL students. Students had an opportunity to learn core concepts, and then build on that concept with the next level text that built on what they had already learned, while also providing new information. Teachers also mentioned that the students enjoyed the books and the small group instruction.

Finally, the teachers provided suggestions for improving the Millmark instructional program. Suggestions included, using some of the black line masters as formative assessments after reading the chapters, making the purple and orange level students' books longer, and adding a lower level of text for students who are limited to non-English speaking. Teachers also requested additional websites, incorporation of more simulations, and use of videos. Other suggestions for improvement related to lesson pacing. One teacher mentioned that she was unable to complete all instruction as planned in a 42-minute period. Another teacher mentioned that during writing, students who worked very slowly often slowed their group.

Conclusions and Recommendations

Students completed an average of 8 lessons with an average of 58 instructional minutes per lesson. Overall, students had a 33% gain in their vocabulary and conceptual comprehension. The mean difference in students' pre- and post-test performance was 5.33, which is a statistically significant gain in student performance attributable to the instruction they received with the *ConceptLinks™ Ecosystems* module.

Teacher survey results indicate that teachers rated the instructional module effective with small group instruction and aligned to their district goals for learning. Teachers also reported that students had high levels of interest and engagement with the books. They reported the differentiated texts as a primary strength of the module. They felt these texts supported learning for all students, but worked particularly well with ELL students. They

stated that the language and text feature were supportive of their students' needs for content area learning. Teachers reported that the books were useful for teaching both content and literacy, especially in the area of content vocabulary. Teachers also reported ease of use with the teacher's guide. The structure of the guide provided effective delivery of instruction and supported classroom management of small group instruction.

There were many differing characteristics among learners and teachers selected to participate in this study. The numbers of grade level readers and struggling readers below grade level were not proportional among the class groups, which makes it difficult to draw conclusions about the impact of the instructional module on various types of learners. Teachers also implemented the lessons in a variety of ways and their minutes of instruction varied widely. These are all variables that impact the limitations of this pilot study. It is recommended that a follow up study be conducted with a control group and structured selection process for matching characteristics of participating teachers and student groups. We recommend conducting site visits, classroom observations, and exit interviews with teachers or teacher groups in any follow up study. Pre-defining characteristics of participants prior to recruiting them for an evaluation study of the instructional modules also is recommended so that the sample size is large enough to substantiate more robust generalizations about specific types of learners in specific types of instructional settings.

More data points are needed with regard to teacher's implementation of the instructional module. The variance among class groups who showed significant gains and other class groups that did not is indiscernible from the available data. The teacher survey data does not provide an explanation for this erratic student performance. For example, Teacher 3 had 4 classes of fifth graders who were all reading at Reading Level 5. Two of these classes showed a mean gain of 2.0 while the other two classes showed zero gain and .83 gain. Teacher 5 also had erratic results. Her 8th graders had a mean gain of -.71 while her 6th graders had a gain of 2.65 and her 7th graders a gain of 3.24. These findings are unexplainable with the current data set.

The effect size for this pilot study is low due in part to the lack of a control group for the study. Expanding the study to include control groups in a structured sample would enhance the ability to draw generalizable conclusions about the efficacy of the instructional module.

The large learning gains found among the regular education students who were reading at grade level during the treatment show the value of providing all types of students with opportunities to learn from the Millmark instruction module. Comprehending non-fiction text can be more challenging for readers than other types of literature. As a result, readers can be challenged to determine importance even if the reading level of the text is at or below their reading level. In other words, students reading at 6th grade level do not necessarily comprehend specific vocabulary or academic concepts unless the instructional program assists with building their content vocabulary and conceptual understanding. While struggling readers and some ELL students in this study benefited more than grade level readers, the emphasis on preparing students to read technical

literature and non-fiction academic text is steadily increasing in America's schools. All students need effective instruction to acquire these higher-order reading and vocabulary development skills.

ELL students in this sample were not necessarily reported as struggling readers in English. For purposes of statistical analysis, a larger sample size of non-ELL and ELL students with special needs is required to verify whether or not the positive results for this study are replicable with a full range of at-risk readers. Expanding the study to include observations and teacher exit interviews would enable the collection of more demographic data on students, and serve to substantiate teacher reported information on surveys. These data could be used to corroborate statistical data. Use of a control group would further support and confirm these data.

Due to students' substantial baseline knowledge in reading visuals and non-fiction text features, evidenced by high scores on these pre-test subscale, we recommend modifying the existing assessment for any follow up studies. Updating these two assessment subscales with more discriminating questions about non-fiction text features and technical graphics or technical pictures would provide a strong measure for ascertaining impact of the instructional module for these purposes.

Professional development was conducted by Millmark consultants. Training for facilitation of the professional development was conducted remotely. This makes it difficult to assess the actual impact and level of understanding among the facilitation participants. Within any follow-up studies, it is recommended that facilitation training be delivered in a face-to-face session, or through web-based video conferencing. Since the studies take place in various school districts across the United States, it is also recommended that data are collected on the structure of the professional development delivered to teachers. Information such as number of participants, setting (before school, after school, during school hours) and location (district training room, classroom, cafeteria) would be helpful to support data analysis of student performance. Finally, collection of data on teacher training is recommended. These data could include evaluation surveys completed after training, as well as questions posed at the exit interview. This would enable more robust data on the relationship between teacher training and student outcomes.

Appendix A: Validation of the Criterion-Referenced Test

Of the 589 students who completed the pre-test, 550 students were used for the pre-test analysis and test validation. There were 12 participating teachers whose students' pre-test data are included in the test validation study. The test was validated for reading levels 4 through 8 only. Eighteen students who were in 4th through 8th grade, who read below Level 4, were considered outliers and removed from this statistical analysis.

The percent of participating students at each reading level is shown in Table 1. The majority of students clustered proportionately across Reading Levels 4-6. Table 2 shows the number of students reading at or below their grade level in grade 4 through 8. There was only one non-ELL student in seventh grade and none in eighth grade. Half of the ELL 7th graders and more than half of the ELL 8th graders read 2.5 levels below grade level and were taught by the same teacher (two class groups).

Table 1. Percent of students at each reading level.

n = 550	Frequency	Percent	Cumulative Percent
Reading Level 4	175	31.8	31.8
Reading Level 5	152	27.6	59.5
Reading Level 6	163	29.6	89.1
Reading Level 7	35	6.4	95.5
Reading Level 8	25	4.5	100.0
Total	550	100.0	

Teachers coded student tests by levels of English language learning and reading level. If a student had been tagged by the school as below grade level in reading, that data was included in the student profile for this study. For purposes of this analysis, students' English language learning status and reading level were combined to create a third code for students - the Student Type variable. There are eight categories in the Student Type variable. There were no students identified as 1.0 or 2.0 below grade level in their reading. The lowest readers were 2.5 below their grade level.

During the spring 2007 pilot of the *ConceptLinks™ Ecosystems* instruction module, the assessment used for measuring students' learning was field-tested for the first time. Field testing is a necessary first step in validating a new test. The item response theory statistics reported in this Appendix are based on the spring 2007 field test of the Millmark assessment, which is a prerequisite to using the test in an experimental or quasi-experimental study with a control group.

The assessment used as a pre- and post-test during instruction with the Millmark Education program was designed to be administered across grades 4 through 8. As a result, student grouping for the analysis is based on reading level rather than grade level (see Student Types in Table 2). Students were classified as ranging in reading ability from Level 4 to Level 8.

Table 2. Shows the results of the Student Type cross tabulation.

Student Type	Grade					Total
	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	
n = 550						
Non-ELL at grade level	84	70	83	1	0	238
Non-ELL 0.5 below grade level	4	7	1	0	0	12
Non-ELL 1.5 below grade level	0	2	2	0	0	4
Non-ELL 2.5 below grade level	0	0	4	0	0	4
ELL at grade level	59	32	36	33	25	185
ELL 0.5 below grade level	3	2	2	1	0	8
ELL 1.5 below grade level	0	1	6	0	0	7
ELL 2.5 below grade level	0	0	18	33	41	92
Total	150	114	152	68	66	550

The 1-Sample Kolmogorov-Smirnov statistical test was used to identify whether or not each Reading Level group was a population with normal distribution. The population of students involved in this pilot study is normal according to the one-sample K-S results in Table 3. The 1-Sample K-S was also conducted for each class group. There were 26 class groups participating in the pilot, some of which had the same teacher. All class groups were found to be normal also.

The normality of the groupings indicates the statistics used to validate the test. In this case, the parametric statistical methods used to analyze the validity of the test included the following statistical steps:

- Determining whether or not the students groups have any outliers.
- Determining the normality of the sample population using the One-Sample Kolmogorov-Smirnov Test. This test was also used for each class group.
- Determining the suitability of the data for factor analysis using the Kaiser-Meyer-Olkin and the Bartlett's test of sphericity.
- Conducting a principal component factor analysis of the vocabulary and comprehension subscales using eigenvalues over 1 for factor selection and the Varimax rotation method with Kaiser Normalization.
- Conducting the Pearson correlation test to verify the strength of the correlations among test items and the total test scores for each subscale: vocabulary and comprehension.
- Conducting the Cronbach Alpha test for reliability.

Table 3. One-Sample Kolmogorov-Smirnov Test of Normality per Reading Level

Total Score		Reading Level 4	Reading Level 5	Reading Level 6	Reading Level 7	Reading Level 8
	N =	175	152	163	35	25
Normal Parameters(a,b)	Mean	17.61	18.90	20.42	16.46	20.36
	Std. Deviation	4.586	5.329	5.317	5.008	2.413
Most Extreme Differences	Absolute	.076	.115	.140	.094	.125
	Positive	.059	.078	.108	.072	.097
	Negative	-.076	-.115	-.140	-.094	-.125
Kolmogorov-Smirnov Z		1.004	1.413	1.793	.558	.623
Asymp. Sig. (2-tailed)		.266	.037	.003	.914	.833

a Test distribution is Normal. b Calculated from pre-test data.

Table 4 shows the results of the Kaiser-Meyer-Olkin and the Bartlett's test of sphericity. These statistics help determine the suitability of the data for factor analysis. These data indicate that factor analysis is an appropriate statistical measure for the Millmark test items.

Table 4. Kaiser-Meyer-Olkin measure of sampling adequacy and the Bartlett's test of sphericity results for the vocabulary and comprehension subscales including the pre-test total score

		Vocabulary Subscale	Comprehension Subscale
Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.768*	.832*
Bartlett's Test of Sphericity	Approx. Chi-Square	857.655	2895.609
	Df	28	171
	Sig.	.000	.000

* >0.5 indicates sampling adequacy

The results of the factor analysis are compiled in the rotated component matrices for the vocabulary subscale, the comprehension subscale, and total test (See Tables 5, 6, and 7 respectively). The initial factor analysis results and the Pearson correlations (see Tables 8 and 9) indicated two test items needed to be eliminated to make the test a stronger measure of students' vocabulary building and comprehension abilities. When a test item is found to be a weak measure, it is removed from the final statistical calculations for the test. Table 10 and 11 show the results from the Cronbach Alpha statistic and Corrected Item-Total Correlations, which indicate that the criterion-referenced test is highly reliable and valid.

Test questions 10 and 20 in the comprehension section were deleted from the data set because the Item Total Correlations among these items and other items in the subscales including the Pre-Test Total were .30 or less which indicates low reliability. The initial factor analysis also showed weak correlations between test questions 10 and 20 and the

other items. The revised factor analysis is included in this report. Test questions 2 and 3 also have weak correlations in the factor analysis, however, these two vocabulary questions have significant correlations with other test items according to the results of the Pearson statistic in Table 8.

Table 5. Factor analysis of the vocabulary subscale.

	Rotated Component Matrix	
	1	2
Pretest Total	.642	.563
Q1	.114	.589
Q2	.044	.666
Q3	.027	.708
Q4	.677	.060
Q5	.822	.084
Q6	.720	.125
Q7	.305	.484

Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

Table 6. Factor analysis of the comprehension subscale.

	Rotated Component Matrix			
	1	2	3	4
Pretest Total	.557	.441	.529	.364
Q8	.186	.675	-.034	.140
Q9	.105	.706	.088	-.010
Q11	.293	.114	.368	-.145
Q12	.165	.628	.083	.068
Q13	.072	.410	.558	-.087
Q14	.069	.561	.180	.078
Q15	.244	.268	.445	.089
Q16	.106	.022	.699	.088
Q17	.172	-.005	.577	.333
Q18	.324	.218	.237	.221
Q19	-.046	.238	.073	.662
Q21	.213	.002	.064	.731
Q22	.312	-.041	.464	.408
Q23	.663	.044	.067	.123
Q24	.674	.159	.175	-.099
Q25	.680	.151	.154	.093
Q26	.614	.160	.154	.062
Q27	.563	.116	.179	.314

Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

Table 7. Factor Analysis of the 25 items in the pre-test.

	Rotated Component Matrix					
	1	2	3	4	5	6
Pretest Total	.568	.390	.437	.362	.353	.278
Q1	.271	.104	.164	-.368	.501	.151
Q2	-.033	.054	.055	.166	.718	-.025
Q3	.027	.053	.033	.058	.727	.073
Q4	.080	.071	.661	.106	.065	-.084
Q5	.067	.046	.771	.113	.076	.104
Q6	.107	.020	.677	.043	.078	.274
Q7	.135	.143	.322	.067	.405	.023
Q8	.133	.706	-.052	.061	.027	.145
Q9	.124	.722	.093	-.015	-.031	.031
Q11	.313	.181	.101	.287	-.050	-.183
Q12	.166	.579	.154	-.022	.163	.063
Q13	.170	.372	.373	.216	.191	-.135
Q14	.067	.598	.020	.185	.085	.013
Q15	.283	.298	.175	.329	.110	.016
Q16	.136	.100	.182	.641	.089	-.046
Q17	.208	.068	.133	.555	.044	.261
Q18	.443	.153	.299	-.053	.030	.261
Q19	.026	.161	.238	-.032	.067	.679
Q21	.150	.028	-.035	.344	.036	.639
Q22	.322	.016	.032	.532	.142	.298
Q23	.631	.010	.044	.120	.087	.077
Q24	.704	.125	.146	.045	.033	-.116
Q25	.667	.131	.089	.136	.065	.085
Q26	.605	.176	.026	.167	-.007	.050
Q27	.511	.161	-.051	.348	.048	.255

Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.

Table 8. Pearson statistics show significantly strong correlations among the vocabulary test items and total pre-test score.

n = 550		Pre-Test Total
Q1	Pearson Correlation	.356(**)
	Sig. (2-tailed)	.01
Q2	Pearson Correlation	.337(**)
	Sig. (2-tailed)	.01
Q3	Pearson Correlation	.349(**)
	Sig. (2-tailed)	.01
Q4	Pearson Correlation	.407(**)
	Sig. (2-tailed)	.01
Q5	Pearson Correlation	.484(**)
	Sig. (2-tailed)	.01
Q6	Pearson Correlation	.478(**)
	Sig. (2-tailed)	.01
Q7	Pearson Correlation	.441(**)
	Sig. (2-tailed)	.01

** Correlation is significant at the 0.01 level (2-tailed).

Table 9. Pearson statistics show significantly strong correlations among the comprehension test items and total pre-test score.

N = 550		Pretest Total
Q8	Pearson Correlation	.393(**)
	Sig. (2-tailed)	.000
Q9	Pearson Correlation	.391(**)
	Sig. (2-tailed)	.000
Q11	Pearson Correlation	.345(**)
	Sig. (2-tailed)	.000
Q12	Pearson Correlation	.445(**)
	Sig. (2-tailed)	.000
Q13	Pearson Correlation	.510(**)
	Sig. (2-tailed)	.000
Q14	Pearson Correlation	.387(**)
	Sig. (2-tailed)	.000
Q15	Pearson Correlation	.510(**)
	Sig. (2-tailed)	.000
Q16	Pearson Correlation	.453(**)
	Sig. (2-tailed)	.000
Q17	Pearson Correlation	.486(**)
	Sig. (2-tailed)	.000
Q18	Pearson Correlation	.508(**)
	Sig. (2-tailed)	.000
Q19	Pearson Correlation	.391(**)
	Sig. (2-tailed)	.000
Q21	Pearson Correlation	.405(**)
	Sig. (2-tailed)	.000
Q22	Pearson Correlation	.522(**)
	Sig. (2-tailed)	.000
Q23	Pearson Correlation	.479(**)
	Sig. (2-tailed)	.000
Q24	Pearson Correlation	.511(**)
	Sig. (2-tailed)	.000
Q25	Pearson Correlation	.560(**)
	Sig. (2-tailed)	.000
Q26	Pearson Correlation	.499(**)
	Sig. (2-tailed)	.000
Q27	Pearson Correlation	.535(**)
	Sig. (2-tailed)	.000

** Correlation is significant at the 0.01 level (2-tailed).

Table 10. Results from the Cronbach Alpha.

	Cronbach's Alpha*	Test Items
Vocabulary Subscale	.729	Questions 1-7 and Pre-Test Total
Comprehension Subscale	.849	Questions 8, 9, 11-19, 21-27 and Pre-Test Total
Total Test		

*> .5 indicates mid-level reliability; >.75 high-level reliability (Hinton, 2004) based on Standardized Items

Table 11. Corrected Item-Total Correlation results for the two subscales and pre-test (25 items).

Vocabulary Subscale Item-Total Statistics - Cronbach		
	Corrected Item-Total Correlation	
Pre-Test Total	.727	
Q1	.360	
Q2	.347	
Q3	.359	
Q4	.414	
Q5	.507	
Q6	.485	
Q7	.440	
Comprehension Subscale Item-Total Statistics – Cronbach		Pre-Test Item-Total Statistics – Cronbach
	Corrected Item-Total Correlation	Corrected Item-Total Correlation
Pre-Test Total	.940	Pretest Total 1.000
Q8	.391	Q1 .312
Q9	.377	Q2 .292
Q11	.332	Q3 .304
Q12	.417	Q4 .368
Q13	.468	Q5 .445
Q14	.371	Q6 .438
Q15	.490	Q7 .402
Q16	.429	Q8 .372
Q17	.471	Q9 .363
Q18	.475	Q11 .312
Q19	.342	Q12 .418
Q21	.382	Q13 .474
Q22	.509	Q14 .356
Q23	.463	Q15 .478
Q24	.498	Q16 .414
Q25	.550	Q17 .451

Q26	.495	Q18	.469
Q27	.537	Q19	.348
		Q21	.362
		Q22	.486
		Q23	.441
		Q24	.475
		Q25	.526
		Q26	.464
		Q27	.503

Appendix B: Sample Population Characteristics

Table 1. Cross tabulation of students per grade level, reading level and type of learner (non-ELL and ELL).

Native English Learners						
GR	Reading Level				Total	
	Reading Level 4	Reading Level 5	Reading Level 6	Reading Level 7		
4	76	0	0	0	76	
5	1	62	0	0	63	
6	4	1	81	0	86	
7	0	0	0	1	1	
Total	81	63	81	1	226	

English Language Learners						
GR	Reading Level					Total
	Reading Level 4	Reading Level 5	Reading Level 6	Reading Level 7	Reading Level 8	
4	55	0	0	0	0	55
5	0	28	0	0	0	28
6	13	6	36	0	0	55
7	0	26	0	32	0	58
8	0	0	37	0	23	60
Total	68	60	73	32	23	256

Table X. Cross tabulation of students categorized by ELL or Non-ELL and reading below their grade level.

		Reading				Total
		At Grade Level	0.5 Below Grade Level	1.5 Below Grade Level	2.5 Below Grade Level	
StudentType	Non-ELL	214	6	2	4	226
	ELL	168	6	6	76	256
Total		382	12	8	80	482

Appendix B: Descriptive statistics for pre- and post-test and gain scores categorized by test subscale

Comprehension Subscale

Table 1. Range of pre- and post-test scores on the comprehension subscale.

Pre-Test Concept Comprehension Subscale			
# of Questions Correct	# Students	Percent	Cumulative Percent
0	7	1.5	1.5
1	24	5.0	6.4
2	39	8.1	14.5
3	40	8.3	22.8
4	49	10.2	33.0
5	45	9.3	42.3
6	74	15.4	57.7
7	102	21.2	78.8
8	102	21.2	100.0
9	0	0	0
Total	482	100.0	
Post-Test Concept Comprehension Subscale			
# of Questions Correct	# of Students	Percent	Cumulative Percent
0	3	.6	.6
1	7	1.5	2.1
2	19	3.9	6.0
3	22	4.6	10.6
4	28	5.8	16.4
5	50	10.4	26.8
6	63	13.1	39.8
7	63	13.1	52.9
8	127	26.3	79.3
9	100	20.7	100.0
Total	482	100.0	

Table 2. Conceptual comprehension subscale Descriptive Statistics - Non-ELL

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	226	0	8	6.04	2.009
Post-Test	226	0	9	7.30	1.943
Gain Score	226	-4	6	1.26	1.733

Table 3. Conceptual comprehension subscale Descriptive Statistics - ELL

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	256	0	8	4.89	2.315
Post-Test	256	0	9	6.09	2.157
Gain Score	256	-4	7	1.20	2.081

Table 4. Conceptual comprehension subscale Descriptive Statistics - all students

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	482	0	8	5.43	2.248
Post-Test	482	0	9	6.66	2.144
Gain Score	482	-4	7	1.23	1.924

Vocabulary Subscale

Table 5. Range of pre- and post-test scores on the vocabulary subscale.

Pre-Test Vocabulary Subscale			
# of Questions Correct	# Students	Percent	Cumulative Percent
Valid -4	2	.4	.4
-3	8	1.7	2.1
-2	22	4.6	6.6
-1	40	8.3	14.9
0	108	22.4	37.3
1	97	20.1	57.5
2	92	19.1	76.6
3	68	14.1	90.7
4	26	5.4	96.1
5	18	3.7	99.8
6	1	.2	100.0
7	0	0	0
Total	482	100.0	

Post-Test Vocabulary Subscale			
# of Questions Correct	# Students	Percent	Cumulative Percent
Valid 0	1	.2	.2
1	14	2.9	3.1
2	22	4.6	7.7
3	44	9.1	16.8
4	70	14.5	31.3
5	80	16.6	47.9
6	81	16.8	64.7
7	170	35.3	100.0
Total	482	100.0	

Table 6. Vocabulary subscale Descriptive Statistics - Non-ELL

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	226	-4	6	1.02	1.711
Post-Test	226	1	7	5.92	1.395
Gain Score	226	1	7	4.90	1.638

Table 7. Vocabulary subscale Descriptive Statistics - ELL

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	256	-4	5	1.32	1.847
Post-Test	256	0	7	4.71	1.774
Gain Score	256	0	7	3.39	1.831

Table 8. Vocabulary subscale Descriptive Statistics - All students

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	482	-4	6	1.18	1.789
Post-Test	482	0	7	5.28	1.716
Gain Score	482	0	7	4.10	1.898

Text Features Subscale

Table 9. Range of pre- and post-test scores on the text features subscale.

Pre-Test Text Features Subscale				
	# of Questions Correct	# Students	Percent	Cumulative Percent
Valid	0	6	1.2	1.2
	1	5	1.0	2.3
	2	19	3.9	6.2
	3	42	8.7	14.9
	4	92	19.1	34.0
	5	125	25.9	60.0
	6	193	40.0	100.0
	Total	482	100.0	
Post-Test Text Features Subscale				
	# of Questions Correct	# Students	Percent	Cumulative Percent
Valid	0	1	.2	.2
	1	8	1.7	1.9
	2	12	2.5	4.4
	3	28	5.8	10.2
	4	60	12.4	22.6
	5	111	23.0	45.6
	6	262	54.4	100.0
	Total	482	100.0	

Table 10. Text features subscale Descriptive Statistics - Non-ELL

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	226	1	6	5.25	1.043
Post-Test	226	1	6	5.52	.875
Gain Score	226	-3	4	.27	.948

Table 11. Text features subscale Descriptive Statistics - ELL

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	256	0	6	4.43	1.415
Post-Test	256	0	6	4.83	1.335
Gain Score	256	-3	6	.40	1.316

Table 12. Text features subscale Descriptive Statistics - All students

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	482	0	6	4.81	1.319
Post-Test	482	0	6	5.15	1.192
Gain Score	482	-3	6	.34	1.159

Visual Literacy Subscale

Table 13. Range of pre- and post-test scores on the visual literacy subscale.

Pre-Test Visual Literacy Subscale				
	# of Questions Correct	# Students	Percent	Cumulative Percent
Valid	0	21	4.4	4.4
	1	59	12.2	16.6
	2	140	29.0	45.6
	3	262	54.4	100.0
	Total	482	100.0	
Post-Test Visual Literacy Subscale				
	# of Questions Correct	# Students	Percent	Cumulative Percent
Valid	0	21	4.4	4.4
	1	51	10.6	14.9
	2	130	27.0	41.9
	3	280	58.1	100.0
	Total	482	100.0	

Table 14. Visual literacy subscale Descriptive Statistics - Non-ELL

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	226	0	3	2.50	.750
Post-Test	226	0	3	2.60	.681
Gain Score	226	-2	3	.09	.650

Table 15. Visual literacy subscale Descriptive Statistics - ELL

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	256	0	3	2.18	.913
Post-Test	256	0	3	2.20	.928
Gain Score	256	-2	3	.02	.918

Table 16. Visual literacy subscale Descriptive Statistics - All students

	N	Minimum	Maximum	Mean	Std. Deviation
Pre-Test	482	0	3	2.33	.855
Post-Test	482	0	3	2.39	.844
Gain Score	482	-2	3	.05	.804