

# The Impact of the Voyager Universal Literacy System<sup>®</sup> in Birmingham City Schools



Birmingham, Alabama

June 2002

Greg Roberts, Ph.D.

Program Evaluation Consultant

Evaluation Research Services

**E R S**

---

Evaluation Research Services

# CONTENTS

---

EXECUTIVE SUMMARY .....	3
OBJECTIVES OF THIS REPORT .....	4
EVALUATION QUESTIONS AND ANSWERS .....	5
SUMMARY OF FINDINGS AND RECOMMENDATIONS ...	11

# EXECUTIVE SUMMARY

---

## THE PROGRAM

The Voyager Universal Literacy System incorporates the five critical components of reading instruction (phonemic awareness, phonics, fluency, vocabulary, and comprehension) as defined in the *No Child Left Behind* Act. Systematic, explicit instruction addresses each of these components, using a structured sequence of skill development.

Voyager Universal Literacy is a comprehensive system, comprised of the following components: in-school comprehensive reading curricula, progress monitoring system, extended-day and summer intervention, home study curriculum, implementation support, and initial and ongoing professional development. The curricula are built upon a detailed scope and sequence. Each day's learning builds upon the previous day's learning. Likewise, skills in each grade level build upon skills and strategies from the previous year.

## THE EVALUATION

Birmingham City Schools, Birmingham, Alabama, implemented the Voyager System during the 2001–2002 school year. Six schools participated in the evaluation. Twelve classrooms were included, with one first grade classroom and one kindergarten classroom selected by the building administrator from each school.

A one-group pretest/posttest design was used. Students took pretests in late October or early November. The Vital Indicators of Progress™ (VIP) and the Woodcock Diagnostic Reading Battery (WDRB) were administered. Posttests were administered in late April and early May. Instruction in each participating

classroom was observed on at least two occasions, once by a Voyager representative and once by a Birmingham district official.

## THE FINDINGS

Data from 177 students were included in the evaluation. Ninety-one students were in kindergarten and 86 were in first grade. Only cases with a pre and a posttest score were used in the evaluation.

Percentile rank (PR) scores on the WDRB indicate that Voyager students made gains beyond those expected of same-age students. Further, number-correct scores on the VIP suggest that more than 60 percent of students mastered the important early literacy skills that predict later success in reading.

When classrooms that failed to adequately implement the program (n=2 kindergarten classrooms) were excluded from the analysis (resulting in a total n of 145, and kindergarten n of 59), the findings were more dramatic. Students increased their average percentile rank by up to 20 points. When spring posttest scores (April/May 2002) for Voyager kindergarten students in implementing classrooms were compared to pretest scores (November 2001) for first grade students before Universal Literacy System implementation, kindergarten students scored significantly higher—66 PR to 39 PR for kindergartners and first graders respectively.

# OBJECTIVES OF THIS REPORT

This report presents findings from evaluation of the Voyager Universal Literacy System™ used in Birmingham, Alabama, during the 2001–2002 school year. The evaluation questions are outlined below. The body of the report is organized around these questions.

What was the Universal Literacy System’s impact on the reading skills of Birmingham first graders and kindergartners participating in the evaluation?

- How did students’ progress compare to early reading benchmarks?\*
- How did students’ progress compare to the progress made by other students?
- What was the relationship of program implementation to student outcomes?

What evidence supports the validity of Vital Indicators of Progress™ (VIP)?

- Does VIP measure important early reading outcomes?\*
- How accurately do these outcomes measure critical reading skills?
- Did Birmingham teachers use VIP correctly?

How well does oral reading fluency predict reading comprehension? How well does story retelling predict reading comprehension?

\*Voyager’s Vital Indicators of Progress™ (VIP) includes standards, or benchmarks, for evaluating students’ progress at different times of the year. Students meeting this standard are considered ‘established’ readers. ‘Emerging’ readers are approaching the benchmark, while ‘struggling’ readers are far short of the benchmark.

For kindergarten students:

- The fall benchmark is at least 10 correct on the one-minute measure of initial sound fluency.
- The end-of-year benchmark is a number-correct of at least 35 on the one-minute measure of phoneme segmentation fluency.

For first grade students:

- The fall benchmark is a number-correct of at least 35 on the one-minute measure of phoneme segmentation fluency.
- The end-of-year benchmark is reading correctly at least 40 words on the one-minute measure of oral reading fluency.

# EVALUATION QUESTIONS AND ANSWERS

---

***What was the impact of the Voyager Universal Literacy System™ on the reading skills of Birmingham first graders and kindergartners participating in the evaluation?***

Impact can be measured in several ways. Two approaches were used in this evaluation. One was a criterion-referenced approach to evaluating student progress. A criterion-referenced approach compares a child's progress to some standard. The standards in this case were the research-based components that characterize early literacy. These components are the foundation for *No Child Left Behind*, recent ESEA legislation on early reading education, as well as the basis for Vital Indicators of Progress™ (VIP).

A norm-referenced approach is a second way to measure student progress. A norm-referenced approach compares a student to other students in the same grade or of similar age. Norm-referenced measurement is the basis for tests used nationally to track student learning. Examples include the California Achievement Test, the Stanford Achievement Test-9, and the Woodcock Diagnostic Reading Battery (WDRB).

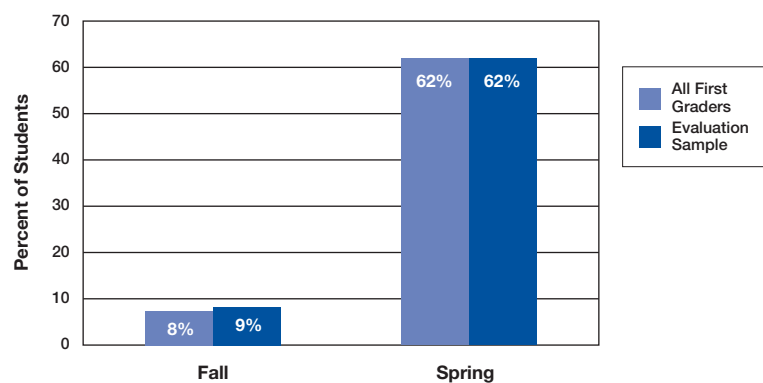
The first two evaluation questions address these two approaches to measuring progress. The first question addresses progress using the VIP benchmarks as criteria. The second question uses normative data from the WDRB to determine students' progress relative to the progress made by the group of similar-age children. Both analyses use data from classrooms that implemented the Universal Literacy System with at least a measurable amount of fidelity. (two classrooms did not implement the program.) The third question in this section addresses the question of implementation and its effect on student progress.

It should be noted also that pretests were administered in late October/early November, meaning that children received about two months of the intervention prior to the evaluation. This exposure to the program may account for the unexpectedly high pretest scores. If so, the student gains presented in this document may underestimate Voyager's impact over a complete school year, a possibility to consider when interpreting the results.

**How did students' progress compare to early reading benchmarks?**

At pretest, about 8 percent of first-grade students in the evaluation sample read 40 or more words per minute correctly on the VIP one-minute oral reading fluency measure. At posttest, 62 percent of students in the evaluation classrooms read 40 or more words. These figures are almost identical to those of the sample of all Birmingham first graders (9 percent and 62 percent; see figure below). This suggests that the first grade evaluation sample may be representative of Birmingham City Schools as a whole, meaning that VIP findings for the group of all first-grade students in Birmingham apply to the group of evaluation-sample students.

Comparison of Woodcock Sample to all Birmingham First Graders

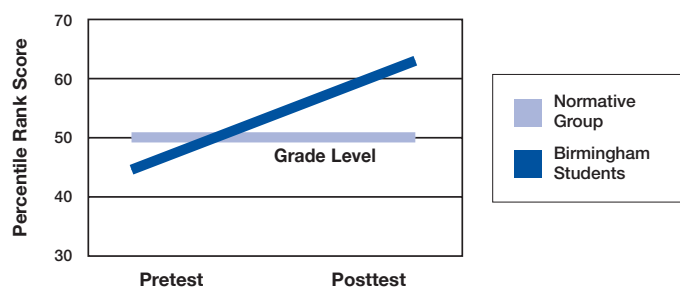


**How did Voyager students' progress compare with children in the normative sample of the Woodcock Johnson Diagnostic Reading Battery?**

The WDRB provides a number of standard scores, including percentile rank (PR) scores. A PR score reflects a child's standing relative to children in the same grade or of similar age. A PR can be interpreted as follows: In a group of 100 randomly selected students of similar age (or in the same grade), this student ranks \_\_\_ in his or her ability to perform the skill being tested. The number in the blank is the PR. A PR of 50 is average. A PR of 99 indicates that a student is in the top 1 percent of the group of students of similar age.

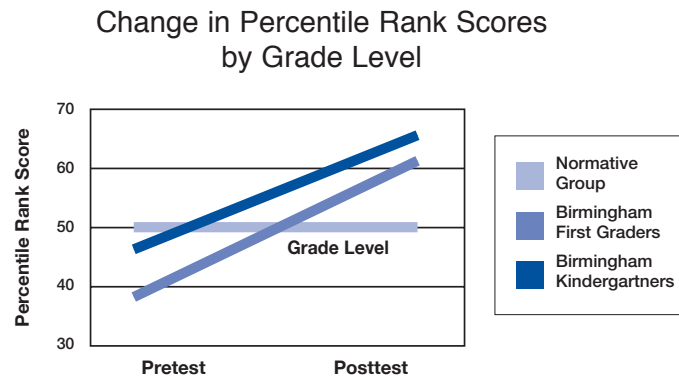
In the figure at right, the group of Voyager students is compared to the normative group on the Broad Reading<sup>1</sup> subtest of the WDRB. The normative group has an average 50 PR at pre and posttest. Voyager students average 45 at pretest and 63 at posttest.

Comparison of Change in Percentile Rank Scores: Birmingham and Normative Group of WDRB

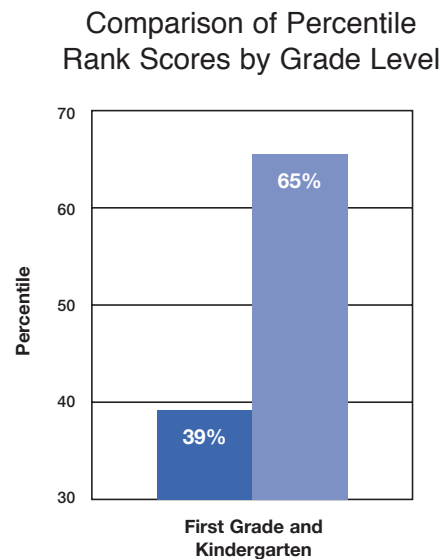


<sup>1</sup>Broad Reading is an aggregate score that includes skills in decoding and text comprehension.

Results of the same analysis for first-graders and kindergarten children are depicted separately in the figure below. First graders moved from an average PR of 39 at pretest to 61 at posttest. Kindergarten students made average gains of 19 PR points, going from an average pretest score of 47 to an average posttest score of 66.



It also may be helpful to compare spring posttest scores (April/May 2002) for Voyager kindergarten students in implementing classrooms to pretest scores (November 2001) for Voyager first grade students (in essence, comparing the initial scores of first grade students before Universal Literacy to the end of year scores of kindergarten students after Universal Literacy). Kindergarten students scored higher than 65 percent of similar-age students at the end of the school year after participating in Universal Literacy, while first graders before participating in Universal Literacy scored the same as or below 39 percent of their peer group, on average. On the WDRB Broad Reading measure, kindergartners in spring (after Universal Literacy) scored an average PR of 66, while first graders in the fall (before Universal Literacy) averaged 39 PR. This means that kindergarten students were 24 PR points higher at the end of the school year, on average, than first graders at the beginning of the school year. This comparison is reflected in the figure at right.



**What was the relationship of program implementation to student outcomes?**

Two observers evaluated program implementation, one observer from Voyager and the other from Birmingham City Schools. The team of two worked together to assign ratings of implementation fidelity to the 12 classrooms participating in the evaluation. A 1 to 10 scale was used, with 1 being *no implementation of the Voyager Universal Literacy System™* and 10 being *total implementation of the Voyager System*. Two kindergarten classrooms scored at or below 3. The remaining classrooms scored 5 and above. One classroom was rated as 5, three classrooms were rated as 6, two as 7, two as 8, one as 9 and one as 10. The average gains in PR by level of implementation are reflected in the following table.

Level of Implementation (Number of Classrooms)	Broad Reading Average PR Pretest	Broad Reading Average PR Posttest
Low (2)	37	39
Middle (6)	30	50
High (4)	61	75

**Classrooms that used the System with a reasonable degree of fidelity had greater success in improving student skills than classrooms that did not use the System.** Interestingly, the average PR for the middle group of classrooms at pretest was considerably lower than that for the high group and somewhat lower than that for the low implementation group. This group also experienced the greatest average gain (20 points). The high group made strong, if somewhat smaller gains, while the low implementation group made virtually no gain in average PR.

It is not surprising that children in the high implementation group began the year well ahead of children from classrooms in the other two groups. More accomplished teachers are likely to teach at schools with more successful students; the same teachers may be more likely to adopt new approaches to teaching. The important finding here relates to the middle implementation group. These data suggest that students who began the year considerably below average and in classrooms using the Voyager System with at least moderate fidelity moved to average and above average standing by the end of the school year. (Students in the high implementation group may have made slightly smaller gains than the middle group due to the effects of regression to the mean.)

**What evidence supports the validity of Vital Indicators of Progress™ (VIP)?**

Evaluators in Birmingham were able to collect data related to the psychometric properties of the VIP. The findings from these data follow.

**Does VIP measure important early reading outcomes?**

The VIP system was developed in collaboration with Dr. Roland Good from the University of Oregon. Dr. Good is the author of the Dynamic Indicators of Basic Early Literacy Skills™ (DIBELS™), which is widely used to monitor early reading progress within a classroom setting. DIBELS and VIP are based on Dr. Good’s ongoing program of research. The benchmarks that serve as foundations of both systems are based on the most current research on early literacy.

**How accurately do the VIP outcomes measure critical reading skills?**

Correlations of VIP data and WDRB data were calculated (table below). Correlation coefficients were high, indicating that skills measured by VIP and WDRB subtests are similar, and that how a student does on the VIP is predictive of how he or she will do on the WDRB.

**Correlation of WDRB and VIP**

	Broad Reading	Basic Reading	Letter-Name Fluency	Phonemic Segmentation fluency	Nonsense Word Fluency	Reading Comprehension of Text
Broad Reading	1.00					
Basic Reading	.91	1.00				
Letter-Name Fluency	.70	.75	1.00			
Phonemic Segmentation Fluency	.64	.63	.61	1.00		
Nonsense Word Fluency	.72	.72	.69	.64	1.00	
Reading Comprehension of Text	.64	.65		.50	.65	1.00

Regression analyses were conducted as well. Pretest WDRB scores and VIP scores accounted for 61 percent of the variance in WDRB basic reading posttest scores and 60 percent of the variance in WDRB Broad Reading posttest scores. Further, when VIP subtests were matched with WDRB subtests, the amount of variance explained increased significantly, suggesting that similarly labeled subtests in the WDRB and the VIP measure comparable constructs. This is strong evidence of VIP’s construct validity.

***Did Birmingham teachers use VIP correctly?***

Two evaluators in Birmingham administered two new oral reading passages as part of a field test of second-grade VIP. Dr. Good trained the evaluators prior to test administration, and inter-reliability analysis indicated that the two evaluators were assigning comparable scores to the same events. Scores collected by the trained evaluators were compared to scores collected by classroom teachers on the existing first-grade VIP oral reading passages. The correlation between teacher-administered first-grade VIP and evaluator-administered field test VIP was .82 for one test passage and .76 for the other test passage, suggesting that the standard procedures and materials used for VIP may provide solid inter-rater reliability for the oral reading fluency part of the first-grade VIP.<sup>2</sup>

***How well do the second-grade oral reading fluency passages predict reading comprehension? How well does story retelling predict reading comprehension?***

The second-grade VIP field test data were used to address these questions. Pretest scores on WDRB Broad Reading and number of words read on the oral fluency measure accounted for 56 percent of the variance in WDRB posttest scores on one test passage and 52 percent on the other. This suggests that VIP oral reading fluency passages are a valid indicator of reading comprehension as measured by the WDRB. However, the addition of story retelling of these passages did not increase predictability of WDRB scores, meaning that story retelling may not contribute to the measurement of reading comprehension (beyond the predictive value of oral reading fluency) at the second-grade level. Further research is recommended.

---

<sup>2</sup> One note: variation due to differences in reading passages was not controlled, so these coefficients may underestimate the actual correlation of teachers' and evaluators' ratings of the oral reading components of VIP.

# SUMMARY AND RECOMMENDATIONS

Participants in the Voyager evaluation in Birmingham, Alabama, made strong gains in early reading skills, as measured by VIP and the WDRB. Further, the validity of the current VIP was supported by data from the WDRB; the two measures yield very comparable results.

Student outcomes are sensitive to levels of program implementation. This is not a surprising finding, given the well-documented effects of inadequate implementation. Voyager’s continuing efforts at providing quality training, ongoing professional development, and institutional support are elements that should be considered by future evaluators.

---

## **About the Program Evaluator**

Greg Roberts holds a Ph.D. in Developmental Educational Psychology and a M.A. in Program Evaluation from The University of Texas at Austin. He is the Assistant Director of the NIH-funded Center for Health Promotion Research in Underserved Populations at The University of Texas at Austin and a member of UT’s general faculty. Dr. Roberts consults independently with a number of educational and health-related organizations. He conducts or has conducted program evaluations for Houghton-Mifflin Company, SERA Learning Incorporated, Kaplan Learning Services, SCORE! Learning Centers, the Centers for Disease Control, the Texas Department of Health, and the Texas Health and Human Services Commission, among others. He has been providing evaluation services to Voyager Expanded Learning since May of 2001.