

Summary of 2009 Striving Readers Projects

Implementation and Evaluation of Targeted Interventions for Struggling Readers after One Year

Submitted by: Abt Associates Inc.

Submitted to: U.S. Department of Education, Office of Elementary and
Secondary Education

Contents

Summary of Striving Readers Projects

- Profile of Illinois' Striving Readers Project and Evaluation
- Profile of Kentucky's Striving Readers Project and Evaluation
- Profile of Louisiana's Striving Readers Project and Evaluation
- Profile of Michigan's Striving Readers Project and Evaluation
- Profile of New York's Striving Readers Project and Evaluation
- Profile of Virginia's Striving Readers Project and Evaluation
- Profile of Washington's Striving Readers Project and Evaluation
- Profile of Wisconsin's Striving Readers Project and Evaluation

Summary of 2009 Striving Readers Projects: Profile of Illinois Department of Education Striving Readers Project and Evaluation¹

Grantee: Illinois Striving Readers Project

Project Director: Sarah McCusker

Local Evaluator: RMC Research Corporation

Principal Investigator: Dimiter Dimitrov

Project Website: http://www.isbe.net/curriculum/reading/html/striving_readers.htm

Setting of the Study

Six high schools from four school districts were selected to participate in the Illinois Striving Readers program. The school districts participating in the program were: Danville Community Consolidated School District #118, Decatur School District #61, Kankakee Public School District #111, and Springfield School District #186. In the 2008-2009 school year, the six participating schools served 2,641 students in the 9th grade. All six schools were Title I eligible. The Striving Readers project was funded for two years, which included one year of implementation of the intervention during the 2010-11 school year.

Intervention Model

Classroom Model as Planned²

Voyager Passport Reading Journeys III (PRJ) is a Cambium Learning Group curriculum. The intervention provides daily, 50-minute lessons with explicit, systematic instruction in critical language skills. (Because of the use of block scheduling in the Decatur School District, the intervention was delivered every other day in a 90-minute block.) PRJ incorporates video segments on DVD, text, and online interactive lessons. It is formatted as a series of ten lessons, or Expeditions, designed to be delivered over the course of one school year on topics related to science, sports and health, culture and diversity, and career development.

Instruction is provided by a trained interventionist, known as a Reading Intervention Teacher (RIT), at each school, with support from Voyager Implementation Specialists. The school division liaisons and project coordinator conduct unscheduled visits to the *Journeys* classrooms to monitor implementation and ensure fidelity to the original design of the model. The Expeditions focus on six instruction practices: (1) explicit vocabulary instruction; (2) direct and explicit comprehension strategy instruction; (3) extended discussion of text meaning and interpretation; (4) instruction in reading foundational skills, such as decoding and fluency; (5)

¹ See final report: Dimitrov, D., Jurich, S., Frye, M., Lammert, J., Sayko, S., Taylor, L. (2012). *Year One Evaluation Report/Impact Study: Illinois Striving Readers*. Arlington, VA: RMC Research Corporation.

² For more information on PRJ, please see the Voyager Passport Reading Journeys III Intervention Profile by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

instruction in writing; and (6) strategies to improve student motivation and engagement in literacy learning.

A library of Lexile-leveled books and magazines on age-appropriate topics is provided for each intervention classroom. The texts, which are primarily nonfiction, have been selected and tested for high interest with high school students. Lessons include whole-group instruction as well as flexible grouping to ensure maximum use of differentiated instructional time.

Professional Development Model as Planned

The professional development model for RITs in the Illinois Striving Readers program consisted of an initial whole group Launch training, self-paced online training, assessment training, and on-site individual monthly support for each teacher. The six RITs (one at each school) were required to attend a statewide data meeting, which focused on how to analyze assessment data and use it to individualize instruction. The RITs also were required to attend the four-day annual convention of the International Reading Association (IRA). Implementation specialists from Voyager delivered the group training.

Over the school year, Voyager specialists planned to provide the RITs with 38 hours of professional development, with 14 hours directly related to the intervention and 24 hours focused on reading research and practice in a group training format. At each school, the principal was held responsible for ensuring that the RITs attended all the required training. In addition, Voyager Implementation Specialists monitored and recorded the participation of the RITs in the training and provided feedback on their progress throughout the year in implementing PRJ in the classroom.

The RITs also had available a maximum of 25 hours of coaching and 16 hours of online modules. Usage of these hours depended on the willingness and need of the interventionist; The coaching, which was provided by the Voyager specialists, involved modeling lessons, observing lessons, and providing feedback. The number of days of coaching that the Voyager specialists provided to each RIT was negotiated beforehand with the school districts. Coaching will consist of modeling lessons, observing lessons, and providing feedback. A minimum of 40 hours of one-on-one, individualized coaching was provided. During their visits, the Implementation Specialists used the Index of Fidelity of Implementation to assess how closely teachers are adhering to the planned classroom model. As part of their supervisory responsibilities, principals were expected to conduct regular in-class visits to observe implementation of PRJ.

Context for Implementation

The Illinois Striving Readers program was a one-year intervention that served struggling readers in grade 9 in the selected schools. Approximately 900 students participated in the study, with 450 students randomly assigned to receive PRJ and 450 randomly assigned to the control group. Random assignment occurred in spring 2010. For those students assigned to the treatment group, PRJ was delivered as a daily, 50-minute supplement to regular English Language Arts classes, replacing an elective class. (For those students in the Decatur School District only), the intervention was delivered every other day in 90-minute blocks.

In the spring of 2010, the six participating schools generated a list of eligible students based on their scores on the 8th grade ACT-EXPLORE Reading assessment. Students who scored at the lowest level - below the 27th percentile - were considered eligible for participation in the Illinois Striving Readers program. For students who do not have ACT-EXPLORE Reading scores, the Illinois Standards Achievement Test was used, and students scoring below the 27th percentile were considered eligible for participation in the program.

Students with IEPs who required more intensive, one-on-one instruction, and those who required highly specialized interventions (i.e., students with intensive cognitive and/or emotional disabilities who would not benefit from group instruction) were excluded from the study. English Language Learners with very limited English skills (Levels 1 and 2) who required basic English language instruction also were excluded from the study.

Evaluation Design

Research Questions

1. What is the effect of one year of exposure to the PRJ supplemental literacy intervention on the reading achievement of struggling readers in 9th grade?
2. What is the fidelity of implementation of PRJ in the study classrooms?

Impact Study

Research Design and Methods: The impact of PRJ on student reading achievement was determined using an experimental research design and within-school random assignment of students. The research question on the impact of PRJ was addressed using a two-level HLM model in which students were nested within schools. Separate analyses were performed for two outcomes measured at the end of grade 9: Gates-MacGinitie Reading Test (GMRT) and ACT-EXPLORE Reading Assessment. Two permanent independent variables in this model were the treatment condition (intervention vs. control) and the pretest measures on the grade 9 GMRT. Additional covariates tested for inclusion in the models were Limited English Proficient status, Special Education status, Free/Reduced Price Lunch eligibility, gender, and ethnicity.

Control Condition: Students assigned to the control condition participated in study hall or elective classes that did not involve explicit reading instruction. They attended their chosen elective during the time when the students in the treatment group were receiving the PRJ supplemental literacy intervention. All students participating in the study attended the same English Language Arts classes as their 9th grade classmates, regardless of assignment to the treatment or control condition. To monitor compliance with these parameters, members of the study evaluation team reviewed the schedule of classes for those students randomly assigned to the control group.

Sample Size: The sample that was randomly assigned included 855 grade 9 students - a treatment group of 427 struggling readers in grade 9 who received PRJ for one year, and 428 struggling readers in the control group, across six schools. For the Gates-MacGinitie Reading Tests-4th Edition (GMRT), the sample for the analysis of impacts of one year of exposure to the

intervention included 232 treatment students and 216 control students in 9th grade. For the ACT-EXPLORE Reading Assessment, the sample for the analysis of impacts included 264 treatment students and 250 control students in 9th grade.

The sample of students was large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .12 on the GMRT after one year of the intervention for grade 9,³ and
- .18 on ACT-EXPLORE Reading Assessment after one year of the intervention for grade 9.

Key Measures of Student Reading Outcomes (Source)

Gates-MacGinitie Reading Tests-4th Edition (GMRT) (External Test Publisher)

ACT-EXPLORE Reading Assessment (Illinois Standards Achievement Test) (State Test)

Implementation Study

Research Design and Methods: Fidelity of implementation was calculated for the PRJ professional development model and the classroom instruction model. For the professional development model, fidelity of implementation was calculated as the number of hours attended relative to the 38 hours required in the logic model. The index was computed by school, since each school had one interventionist. For example, if an interventionist attended 30 of the 38 required hours of professional development, the interventionist/school score was 0.79. The evaluators established a score of 1.00 or above as representing adequate fidelity to the professional development model, while scores below 1.00 were considered inadequate. (An RIT could achieve a score higher than 1 if they attended more training than was required.)

For the classroom instruction model, fidelity of implementation was rated based on two observations of each RIT. Four aspects of classroom instruction were rated: amount of instruction, differentiation of instruction, classroom management, and use of assessments. The four factors were weighted and summed to 100 percent. Fidelity of implementation of the classroom model was defined as: low = 0-69 percent; medium = 70-89 percent; high = 90-100 percent.

Evaluation Findings

Fidelity of Implementation of the Intervention Model

For fidelity of implementation of the PRJ professional development model, all interventionists received an index of 1.0 or above, which was defined as adequate implementation. Based on the delivery of the classroom model observed during site visits, two of the six interventionists were classified as attaining high fidelity of implementation (score at or above 0.90), three were classified as attaining medium fidelity (scores between 0.70 and 0.89), and one was classified as attaining low fidelity (score below 0.70).

³ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

Impact of the Intervention on Student Reading Outcomes

There were no statistically significant impacts on the reading achievement of struggling readers in 9th grade after one year of exposure to PRJ, with effect sizes of 0.02 on the GMRT and -0.09 on the ACT-EXPLORE Reading assessment.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Targeted Intervention

Strengths:

- The analysis of the impact of one year of PRJ on reading achievement (GMRT and ACT-EXPLORE) meets What Works Clearinghouse (WWC) evidence standards. Analyses that meet WWC evidence standards make us the most confident that the effect we see is due solely to the intervention alone, and not to the many other factors that are at play in schools and in the lives of students, such as teachers, school and family.
- Eligibility for random assignment was determined systematically, using a predetermined cutoff score on a test of reading achievement (the ACT-EXPLORE Reading assessment).
- Random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.
- There was no evidence that other factors (e.g., other reading programs or district policies) were implemented in ways that would have undermined the evaluators' ability to attribute impacts to Voyager Passport Reading Journeys III (PRJ).
- The evaluation employed two reading tests as outcome measures. The Gates MacGinitie (GMRT) 4th edition measures reading comprehension, and EXPLORE, a component of the ACT testing system, measures academic achievement in English, mathematics, reading and science. There was no reason to believe that students assigned to the treatment group had more experience taking the tests than did the control group students, or that the tests measured skills specific to the intervention, both of which could have undermined confidence in the impact estimates.
- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement was included in the models to increase the precision of the impact estimates.

- While some students were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggested that the integrity of the original randomized design was preserved, and that treatment and control groups continued to be statistically equivalent on all measured and unmeasured characteristics at follow-up.
 - 39.9 percent of grade 9 students who received one year of PRJ were unable to participate in follow-up data collection on the EXPLORE. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 3.4 percent. This amount of attrition is within the acceptable range established by WWC standards.⁴
 - 47.6 percent of grade 9 students who received one year of PRJ were unable to participate in follow-up data collection on the GMRT. The level of attrition did not differ substantially across the treatment and control groups; differential attrition rate was 3.9 percent. This amount of attrition is within the acceptable range established by WWC standards.³

Weaknesses:

- None.

⁴ For more information, please see Appendix A-Assessing Attrition Bias, of the WWC Procedures and Standards Handbook, available at: <http://ies.ed.gov/ncee/wwc/help/idocviewer/Doc.aspx?docId=19&tocId=7>.

Summary of 2009 Striving Readers Projects: Profile of Kentucky Department of Education Striving Readers Project and Evaluation¹

Grantee: Kentucky Department of Education

Project Director: Cindy Parker

Local Evaluator: Collaborative Center for Literacy Development (CCLD)

Principal Investigator: Susan Chambers Cantrell

Project Website: <http://www.kentuckyliteracy.org/sr>

Setting of the Study

The Kentucky Striving Readers program selected nine high schools in nine school districts in northern and southeastern Kentucky that met eligibility requirements and were underserved by other reading partnerships. The districts were: Covington-Holmes, Jackson County, Pendleton County, Lincoln County, George Rogers Clark, Lewis County, Knox Central, Clay County, and Garrard County. The average percentage of students receiving free/reduced price meals at the participating high schools at the time of the study was 62 percent. For the 2008-2009 school year, these schools served 2,415 grade 9 students. Two of the nine participating high schools made Adequate Yearly Progress in the 2008-2009 school year, while six did not, and one was not reported. None of the participating high schools or their feeder middle schools had an average score at or above the college readiness benchmarks for the ACT or EXPLORE test in reading. The Striving Readers project was funded for two years, which included one year of implementation of the intervention during the 2010-11 school year.

Intervention Model

Classroom Model as Planned²

Kentucky's Striving Readers program implemented the Kentucky Cognitive Literacy Model (KCLM) intervention. KCLM was developed by the Kentucky Department of Education literacy staff and is designed as a one-year supplemental literacy intervention for eligible 9th grade students at participating high schools. Students randomly assigned to the intervention meet for at least 50 minutes every day during the entire school year. Each intervention classroom has between 15 and 20 students.

The KCLM curriculum emphasizes the need for automatic cognitive processing and strong literacy skills. In each lesson, students are expected to engage in activities that develop their reading, study, reflective, and communication skills. In the KCLM curriculum, reading, writing,

¹ Cantrell, S.C., Carter, J.C., Rintamaa, M. (2012). *Striving Readers Cohort II Evaluation Report: Kentucky*. (2012). Lexington, KY: Collaborative Center for Literacy Development.

² For more information on *KCLM*, please see the Kentucky Cognitive Literacy Model Intervention Profile by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

and discussion are explicitly linked to one another and embedded in daily learning opportunities. In addition, students are assumed to be motivated through high-interest reading materials, writing and discussion activities, project-based learning, and collaborative projects. Students are also taught how to employ effective comprehension and strategic processing strategies.

Teachers of intervention classes (“interventionists”) were expected to set regular goals and actively monitor student progress. Students kept interactive journals for frequent review by the intervention teacher. Classroom logs and checklists are kept on group discussions and dynamics. Assessments in fluency and decoding are administered frequently for students exhibiting difficulties with oral fluency.

Professional Development Model as Planned

Literacy consultants from the Kentucky Department of Education (KDE) trained the intervention teachers and provided additional professional development and support over the course of the year. A five-day Summer Training Academy was held in July of 2010. During this initial group training, interventionists were introduced to essential instructional practices and were familiarized with the structure and content of the KCLM intervention. The group training was planned to prepare the interventionists to implement a fully functioning intervention class beginning in fall 2010.

In addition to the initial group training, interventionists received one-on-one coaching and support from the literacy consultants at KDE. Every two to three weeks, a professional development trainer visited each of the participating nine schools to provide in-class support. Each interventionist was expected to receive a minimum of 16 hours of individualized coaching and feedback sessions. Interventionists also were expected to participate in at least 50 hours of online professional learning, including book studies, online courses, and webinars for a total of at least 66 hours of individual support per teacher.

Context for Implementation

Approximately 900 grade 9 struggling readers participated in the study, with half of the students randomly assigned to receive the intervention and half to the control group. Random assignment took place in the spring before the implementation year. In the spring of 2010, all 8th grade students were given the Group Reading Assessment and Diagnostic Evaluation (GRADE). Students whose test scores were equivalent to two or more years below grade level—indicated by a score of less than 40—were eligible for the Kentucky Striving Readers program. Evaluators scored the assessment and created a list of eligible students at each school, to be randomly assigned to either the treatment or control group. Special education students who were assigned to a resource room for the entire school day were excluded from participation in the study. For those students assigned to the treatment group, the KCLM intervention was implemented as a daily, 50 minute supplement to regular English Language Arts classes, replacing an elective class.

Evaluation Design

Research Questions

1. What is the impact of one year of exposure to KCLM supplemental reading intervention on the reading achievement of struggling readers in grade 9?
2. What is the impact of one year of exposure to KCLM supplemental reading intervention on the writing achievement of struggling readers in grade 9?
3. What is the impact of one year of exposure to KCLM supplemental reading intervention on the reading strategy use of struggling readers in grade 9?
4. What is the impact of one year of exposure to KCLM supplemental reading intervention on the reading motivation of struggling readers in grade 9?
5. What is the fidelity of implementation of KCLM in the study classrooms?

Impact Study

Research Design and Methods: The intervention's impact on student reading achievement was tested using an experimental research design and within-school random assignment of students. The impact of the KCLM intervention on student achievement in reading and writing, motivation, and reading strategies outcomes was evaluated using a two-level HLM design. At the student level, the spring outcome variable (achievement, strategy use, or motivation) was modeled as a function of the fall outcome variables entered as a covariate (centered by school), intervention/control status and four student-level demographic variables: gender, ethnicity, free/reduced lunch status, and special education.

Control Condition: Students randomly assigned to the control group could select an elective class as part of their regular 9th grade schedule. A wide range of electives were available, including band, chorus, civics, and physical education. Students assigned to the treatment or the control condition attended regular English Language Arts classes along with their 9th grade classmates not participating in the study.

Sample Size: The samples that were randomly assigned included 319 struggling readers in 9th grade who were assigned to receive the KCLM intervention for one year, and 318 struggling readers assigned to the control group, across 9 schools. For the Group Reading Assessment and Diagnostic Evaluation (GRADE), the sample for the analysis of the impact of one year of exposure included 232 treatment students and 253 control students in 9th grade. For the Kentucky Writing Assessment, the sample for the analysis of the impact of one year of exposure included 172 treatment students and 193 control students in 9th grade. For the Metacognitive Awareness of Reading Strategies Inventory (MARSI) and Motivation for Reading Questionnaire scores, the sample for the analysis of the impact of one year of exposure included 154 treatment students and 173 control students in 9th grade.

The sample of students for the impact analysis is large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .21 on the GRADE after one year of the intervention for 9th grade,³
- .26 on the Kentucky Writing Assessment after one year of the intervention for 9th grade,
- .28 on the Metacognitive Awareness of Reading Strategies Inventory (MARSIS) after one year of the intervention for 9th grade, and
- .25 on the Motivation to Read Questionnaire after one year of the intervention for 9th grade.

Key Measures of Student Reading Outcomes (Source)

Group Reading Assessment and Diagnostic Evaluation (GRADE) (External Test Publisher)
 Kentucky Writing Assessment (State Test)
 Metacognitive Awareness of Reading Strategies Inventory (MARSIS) (Mokhtari & Reichard)
 Motivation for Reading Questionnaire (Wigfield & Guthrie)

Implementation Study

Research Design and Methods: The fidelity of implementation of the professional development included a score for number of days (out of five) that each interventionist attended the summer training model. In addition to summer training, interventionists were assessed in terms of the amount of in-person coaching they received from KDE staff throughout the year, calculated as number of visits and total hours of coaching across these visits. Level of fidelity of participation in the summer training was defined as adequate if interventionists attended at least three of the five days. Adequate amount of in-person coaching was defined as a minimum of four ite visits and 24 hours of coaching.

To assess the fidelity of implementation of the classroom model, each interventionist was observed twice during the year, and their instructional practices were scored using a standardized observation protocol for KCLM classes. The observation protocol included two components: a checklist of essential KCLM components and a quality rubric for assessing teachers’ implementation quality. It was expected that teachers would implement some aspect of each of four model components (motivation and engagement, strategic processes, instructional strategies, and communication) during each class period but it was not expected that teachers would implement every aspect of each component each class period. The protocol yielded information on which aspects of each component were implemented during observations and whether the quality of implementation was (a) developing, (b) adequate, or (c) exemplary. Percentages of observations that included each component were computed, and the proportion of observations rated at each quality level was calculated.

Evaluation Findings

Fidelity of Implementation of the Intervention Model

Seven of the nine interventionists attended all five days of the summer training. One interventionist attended four days and was absent one day due to illness. This material was

³ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80 percent power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

covered with the absentee during the following day's training during lunch and breaks. Another interventionist who did not attend the summer training was trained one-on-one during three days at the interventionist's school. During the school year there were three regular training dates. All interventionists attended two of these dates, and eight attended the third. The absent interventionist made up the missed training date via Skype with KDE staff. Interventionists also attended the Kentucky Reading Association conference. Overall, participation in the professional development the KCLM teachers and administrators was high, with 100 percent of teachers and administrators participating fully in the training either through the formal training dates or through makeup training.

Based on the observations of the treatment classroom, teachers were implementing only some of the components with high fidelity. Overall, the intervention was rated as being implemented with adequate quality in three of the nine classrooms.

Impact of the Intervention on Student Reading Outcomes

There were no statistically significant impacts on the reading or writing achievement of struggling readers in grade 9 after one year of exposure to KCLM, with effect sizes of -0.06 on the GRADE and $.07$ on the Kentucky State Writing Assessment.

There were statistically significant impacts on the use of reading strategies and on reading motivation of struggling readers in grade 9 after one year of exposure to KCLM, with an effect size of $.25$ on the MARSII and $.22$ on the Motivation for Reading Questionnaire.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Intervention

Strengths:

- The analysis of the impact of one year of KCLM on reading achievement (GRADE) meets What Works Clearinghouse (WWC) evidence standards. Analyses that meet WWC evidence standards make us the most confident that the effect we see is due solely to the intervention alone, and not to the many other factors that are at play in schools and in the lives of students, such as teachers, school and family. Abt did not examine whether the analyses of the impact of one year of KCLM on other outcomes meets WWC standards because the review was restricted to measures of reading achievement.
- Eligibility for random assignment was determined systematically, using a predetermined cutoff score on a test of reading achievement (the GRADE).
- Random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.
- There was no evidence that other factors (e.g., other reading programs or district policies) were implemented in ways that would have undermined the evaluators' ability to attribute impacts to the Kentucky Cognitive Literacy Model.

- The evaluation employed one reading test as an outcome measure. The Group Reading Assessment and Diagnostic Evaluation (GRADE). There was no reason to believe that students assigned to the treatment group had more experience taking the tests than did the control group students, or that the tests measured skills specific to the intervention, both of which could have undermined confidence in the impact estimates.
- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement was included in the models to increase the precision of the impact estimates.
- While some students were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggested that the integrity of the original randomized design was preserved, and that treatment and control groups continued to be statistically equivalent on all measured and unmeasured characteristics at follow-up.
 - 23.9 percent of grade 9 students who received one year of the Kentucky Cognitive Literacy Model were unable to participate in follow-up data collection on the GRADE. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 6.8 percent. This amount of attrition is within the acceptable range established by WWC standards.⁴

Weaknesses:

- The KCLM Program was developed by the Kentucky Department of Education literacy staff to specifically address the needs of struggling adolescent readers in Kentucky. Therefore, the model is not readily or commercially available to other states or districts who might wish to replicate it.

⁴ For more information, please see Appendix A-Assessing Attrition Bias, of the WWC Procedures and Standards Handbook, available at: <http://ies.ed.gov/ncee/wwc/help/idocviewer/Doc.aspx?docId=19&tocId=7>.

Summary of 2009 Striving Readers Projects: Profile of Louisiana Department of Education Striving Readers Project and Evaluation¹

Grantee: Louisiana Department of Education

Project Director: Jill B. Slack

Local Evaluator: SEDL

Principal Investigator: Michael Vaden-Kiernan

Project Website: http://www.sedl.org/new/pressrelease/20091130_212.html

Setting of the Study

Ten schools in four parishes (or school districts) participated in the Louisiana Striving Readers program. In the 2008-2009 school year, these schools served 3,195 students in grades 6 and 7. All participating schools were Title I eligible, and in 2009, the percentage of students at these schools receiving free/reduced price lunch ranged from 55 to 92 percent. In the 2008-2009 academic year, an average of 41 percent of grade 6 students and 45 percent of grade 7 students scored “*Below Basic*” on the Louisiana state test (Integrated Louisiana Assessment Program or *iLEAP*). The Striving Readers project was funded for two years, which included one year of implementation of the intervention during the 2010-11 school year.

Intervention Model

Classroom Model as Planned²

Voyager Passport Reading Journeys (PRJ) is a Cambium Learning Group curriculum. The intervention provides daily, 50-minute lessons with explicit, systematic instruction in critical language skills. The program incorporates video segments on DVD, text, and online interactive lessons. It is formatted as a series of lessons, or Expeditions, designed to be delivered over the course of one school year on topics related to science, social studies, and literature. PRJ contains different levels for different grades. Each level is designed for one full academic year with 15 two-week Expeditions with optional add-ons (reteach or writing). While each level teaches similar strategies and reinforces common skills, there are unique topics and Expedition content that is age/grade appropriate and of increasing complexity for higher grades. The intervention is called PRJ Beginnings for 6th graders and PRJ I for 7th graders.

The Expeditions focus on six instruction practices: (1) explicit vocabulary instruction; (2) direct and explicit comprehension strategy instruction; (3) extended discussion of text meaning and

¹ See final report: Vaden-Kiernan, M., Caverly, S., Bell, N., Sullivan, K., Fong, C., Atwood, K., Borman, G., Park, S.J., Hughes, D. (2012). *Louisiana Striving Readers. Final Evaluation Report*. Austin, TX: SEDL Research and Evaluation.

² For more information on *PRJ*, please see the Voyager Passport Reading Journeys Intervention Profile by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

interpretation; (4) instruction in reading foundational skills, such as decoding and fluency; (5) instruction in writing; and (6) strategies to improve student motivation and engagement in literacy learning.

A library of Lexile-leveled books and magazines on age-appropriate topics is provided for each intervention classroom. The primarily nonfiction texts were selected and tested for high interest with middle and high school students. Lessons include whole-group instruction as well as flexible grouping to ensure maximum use of differentiated instructional time.

Professional Development Model as Planned

For Striving Readers, the curriculum was implemented by a trained interventionist at each school, with support from Voyager Implementation Specialists. The professional development model for the ten intervention teachers in the Voyager program consisted of an initial group training, online self-paced training, on-site individual monthly support, and a cohort meeting. Implementation specialists from Voyager delivered the training and ongoing support for the interventionists.

Starting in August 2010, each interventionist received 40 hours of professional development in a group training format. These 40 hours consisted of a 2-day “launch” training, which included training in the Strategic Online Learning Opportunities (SOLO) online system, a cohort meeting, and online product training. District coordinators and principals were encouraged to attend the initial launch training. Principals also were required to attend an additional 8 hour meeting.

The in-class coaching was intended to address individual needs of the interventionists. The Voyager Implementation Specialists made nine monthly in-class coaching visits to the classrooms, with each visit lasting approximately four hours. In these sessions, the Specialists conducted classroom observations and provided feedback, lesson modeling, and side-by-side coaching. The specialists also completed the Fidelity of Implementation Checklist and met with the principal of the school during each visit to discuss how the intervention was progressing. In addition, a representative from the Louisiana Department of Education visited each participating school once a month, and the principal of each school made semi-annual visits to each classroom, meeting with teachers to discuss start-up needs and ensuring that teachers are adhering to the program components.

Context for Implementation

The Louisiana Striving Readers program served 6th and 7th grade students in the selected schools. Each level of PRJ - Beginnings and Journeys I - was offered as a one-year supplemental intervention program. Participation in Journeys I does not require completion of Beginnings. For Striving Readers, the Journeys intervention curriculum was implemented as a supplement to regular English Language Arts instruction, replacing an elective class and adding 50 additional minutes of reading instruction each school day for students in the treatment group. PRJ was implemented by literacy intervention teachers that were hired by the school districts (based on selection criteria and requirements established by LDOE) to provide the supplemental program activities during an elective period devoted to this offering at each school.

Approximately 1437 grade 6 and 730 grade 7 student were eligible for the study. Eligibility was based on students' scores on the prior year's *iLEAP*: student who scored "Below Basic" on the test were eligible. Half of these students were randomly assigned to the treatment group and half to the control group, beginning in the spring of 2010. (Students excluded from participation in the study included special education students who did not have an *iLEAP* score because they took an alternative state test and students for whom the PRJ curriculum did not meet the English Language Arts needs of their IEP).

Evaluation Design

Research Questions

1. What is the impact of one year of exposure to the PRJ supplemental literacy program on reading achievement of struggling readers in grades 6-7?
2. What is the impact of one year of exposure to the PRJ supplemental literacy program on reading motivation of struggling readers in grades 6-7?
3. What is the fidelity of implementation of PRJ in the study classrooms?

Impact Study

Research Design and Methods: The impact of PRJ on student reading achievement was determined using an experimental research design with random assignment of students within schools. The impact of PRJ was tested using a two-level hierarchical linear model (HLM) with pretests as covariates and spring posttests on the Group Reading Assessment and Diagnostic Evaluation (GRADE), the *iLEAP*, and the Motivation for Reading Questionnaire as the dependent variables.

Control Condition: All students in the control group attended regular English Language Arts classes along with their classmates in the treatment group and students not participating in the study, but did not participate in PRJ or any other reading-focused supplemental intervention. Therefore, both treatment and control students received the same instructional support provided during the regular English classes, which focused on strategies for comprehension and writing. Students assigned to the control group attended a chosen elective while the students in the treatment group received the PRJ supplemental literacy intervention. Control students were able to select from a variety of elective classes that did not include direct literacy instruction.

Sample Size: The samples that were randomly assigned included a treatment group of 616 struggling readers in grades 6-7 who received PRJ for one year, and 610 struggling readers in the control group, across 10 schools. For both *iLEAP* subtests, the sample for the impact analysis of one year of the intervention included 548 treatment students and 554 control students in grades 6-7. For the Group Reading Assessment and Diagnostic Evaluation (GRADE) Reading Comprehension and Overall Reading subtests, the sample for the impact analysis of one year of the intervention included 455 treatment students and 489 control students in grades 6-7; for the GRADE Vocabulary subtest, there were 454 treatment students and 487 control students in grades 6-7. For the reading motivation score, the sample for the impact analysis of one year of the intervention included 484 treatment students and 493 control students in grades 6-7.

The sample of students for the impact analysis is large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .15 on the iLEAP English Language Arts after one year of the intervention for grades 6-7,³
- .16 on the iLEAP Reading after one year of the intervention for grades 6-7,
- .16 on the GRADE Comprehension and Vocabulary subtests after one year of the intervention for grades 6-7,
- .15 on the GRADE Overall Reading test after one year of the intervention for grades 6-7, and
- .14 on the Motivation to Read Questionnaire after one year of the intervention for grades 6-7.

Key Measures of Student Reading Outcomes (Source):

Integrated Louisiana Educational Assessment Program (*iLEAP*): English Language Arts, Reading (State Test)

Group Reading Assessment and Diagnostic Evaluation (GRADE): Reading Comprehension, Vocabulary, Overall Reading (External Test Publisher)

Motivation for Reading Questionnaire (Wigfield & Guthrie, 1997)

Implementation Study

Research Design and Methods: The fidelity of implementation of the intervention model was studied in the 10 treatment schools. The professional development model included five formats for intervention teachers, delivered and/or facilitated by Louisiana State Department of Education (LDOE) staff and Voyager implementation specialists: Launch Training, Online Product Training, Ongoing Consultative Support, Cohort Meetings, and Principal Training. LDOE staff planned to visit each school at least eight times during the school year to provide technical assistance during and in between Voyager implementation specialist visits to ensure program fidelity. District coordinators planned to provide additional support by coordinating project efforts at the local level and serving as liaisons between and among partners. Principals also planned to visit each intervention teacher once a week during the initial month of implementation and then twice a month over the remaining school year. In total, the project planned to provide a minimum of 86 hours of professional development to teachers over the school year. District and LDOE staff would receive a total of 30 and 18 hours of professional development, respectively, and principals, 36 hours across the planning year and first school year of implementation. Level of fidelity of implementation of the professional development model was defined as: 0 = 51 or fewer hours (low); 1 = 52 to 68 hours (medium); and 2 = 69 or more hours (high). Adequate amount of professional development was defined as having a high level of implementation model or having received 80 percent or more of the proposed professional development hours.

To assess the fidelity of implementation of the classroom model, each intervention teacher was observed multiple times. For each observation, the teacher was rated on a scale from 1 to 4 points on 11 items that aligned with the key instructional practices of PRJ.⁴ The scores assigned

³ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

⁴ The components include Guide, Components, Strategies, Feedback, Pacing, Involvement, Monitoring, Grouping, Tasks, Routines, and Management.

by observers to each of the 11 items were summed for a “score total” and averaged for a “score average” for each classroom section observed.

The fidelity scores for the professional development and the classroom models were combined to create a more inclusive version of the index. The combined index was drawn from multiple sources of data collected over the span of the study, including the amount of professional development delivered, ongoing support provided by the developer’s coaches, and observation scores of teachers’ program implementation. Scores used for the professional development portion of the combined FOI index were derived from the more extensive professional development model discussed earlier. However, only those components of the model specified by the developer as the training and support necessary for implementation of the PRJ program (i.e., not the SR grant) were included in the PRJ combined index. Professional development components included in the combined index were: a composite score for training (combined Launch, Cohort, and online trainings) and a total score for amount of Voyager coaching hours received by teachers. Professional development scores were based on the number of hours teachers accumulated in each category discussed above, and were assigned a 0-2 total value, according to the low, medium, and high adequacy ratings established for overall professional development (0-0.59=low adequacy or 0; 0.60-0.79=medium adequacy or 1; and 0.80-1.0=high adequacy or 2). The classroom components included the total score for instruction, described above, as well as scores for adherence and process factors for the classroom implementation model. A total index score range of 0 to 10 points was possible for each teacher, representing a cumulative measure for the combined fidelity index.

Evaluation Findings

Fidelity of Implementation of the Intervention Model

For the professional development model, only one of the ten teachers reached a medium level of implementation for a fidelity score that combined amount of training and coaching support. All of the teachers were below adequate, meaning they received less than 80 percent of the planned professional development identified in the definition of fidelity of implementation of the professional development model. When the types of professional development are separated, 10 of the 11 intervention teachers received adequate levels of professional development, while none of the teachers received an adequate level of coaching support. While teachers were the primary targets of professional development, within the implementation model proposed by Louisiana Striving Readers, project key members of state, district, and school staff also received professional development to support their teacher’s implementation of the PRJ program. An implementation index was also determined for the amount of professional development that principals and district representatives attended. For the four district administrators, three administrators participated at a high level and one at a medium level. For the ten school administrators, 50 percent participated at a high level, 30 percent at a medium level, and 20 percent at a low level.

The fidelity of implementation of the classroom model was at an adequate across all teachers/classrooms, with the majority of teachers implementing with high levels of fidelity to the PRJ classroom model. The average score for implementation of the classroom model was 3.4

out of 4. Across the 10 intervention teachers, nine (90 percent) had an average score of 3.0 or higher.

Additional results indicated that relative to a more “typical” PRJ model (combined professional development and classroom models), the intervention was implemented with medium to high levels of adequacy across the 10 schools in the study.

Voyager defined the pacing schedule for the implementation of PRJ and specified that the completion of 15 Expeditions (containing 10 lessons each) constituted a "full dose" of the intervention. The average number of Expeditions completed by the intervention teachers was 11.4, with a range from seven to 15 Expeditions completed across the ten teachers.

The combined index scores indicated that 50 percent of teachers demonstrated 80 percent (8 out of 10) or higher levels of implementation, 20 percent demonstrated 70 percent (7 out of 10) levels of implementation, and 30 percent demonstrated 50 percent levels of implementation. These results indicate that relative to a more “typical” PRJ combined (professional development and classroom) implementation model, the program was implemented with medium to high levels of adequacy across the 10 schools in the study.

Impact of the Intervention on Student Reading Outcomes

There was a statistically significant impact on the reading achievement of struggling readers in grades 6-7 after one year of exposure to PRJ, as measured by the GRADE. There were significant impacts on the GRADE overall reading score, vocabulary, and reading comprehension with effect sizes of .27, .13 and .30, respectively.

There were no statistically significant impacts on the reading achievement of struggling readers in grades 6-7 after one year of exposure to PRJ, as measured by the *i*LEAP, with effect sizes of .06 on the overall ELA score for the *i*LEAP and -.005 on the reading subscale for the *i*LEAP.

There was no statically significant impact on the reading motivation of struggling readers in grades 6-7 after one year of exposure to PRJ, with an effect size of .06 on the total score for the Motivation for Reading Questionnaire.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Intervention

Strengths:

- The analysis of the impact of one year of PRJ on reading achievement (GRADE and *i*LEAP) meets What Works Clearinghouse (WWC) evidence standards. Analyses that meet WWC evidence standards make us the most confident that the effect we see is due solely to the intervention alone, and not to the many other factors that are at play in schools and in the lives of students, such as teachers, school and family. Abt did not examine whether the analysis of the impact of one year of PRJ on student motivation meets WWC standards because the review was restricted to measures of reading achievement.

- Eligibility for random assignment was determined systematically, using a predetermined cutoff score on a test of reading achievement (the Integrated Louisiana Educational Assessment Program Reading test).
- Random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.
- There was no evidence that other factors (e.g., other reading programs or district policies) were implemented in ways that would have undermined the evaluators' ability to attribute impacts to Voyager Passport Reading Journeys (PRJ).
- The evaluation employed two reading tests as outcome measures. The Group Reading Assessment and Diagnostic Evaluation (GRADE) Reading Comprehension, Vocabulary, and Oral Language subtests, and the *i*LEAP, the Louisiana state English Language Arts test. There was no reason to believe that students assigned to the treatment group had more experience taking the tests than did the control group students, or that the tests measured skills specific to the intervention, both of which could have undermined confidence in the impact estimates.
- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement was included in the models to increase the precision of the impact estimates.
- While some students were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggested that the integrity of the original randomized design was preserved, and that treatment and control groups continued to be statistically equivalent on all measured and unmeasured characteristics at follow-up.
 - 10.1 percent of grade 6-7 students who received one year of PRJ were unable to participate in follow-up data collection on the *i*LEAP English Language Arts and Reading tests. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.9 percent. This amount of attrition is within the acceptable range established by WWC standards.⁵
 - 23.0 percent of grade 6-7 students who received one year of PRJ were unable to participate in follow-up data collection on the GRADE Reading Comprehension subtest. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 6.3 percent. This amount of attrition is within the acceptable range established by WWC standards.
 - 23.2 percent of grade 6-7 students who received one year of PRJ were unable to participate in follow-up data collection on the GRADE Vocabulary subtest. The level of attrition did not differ substantially across the treatment and control

⁵ For more information, please see Appendix A-Assessing Attrition Bias, of the WWC Procedures and Standards Handbook, available at: <http://ies.ed.gov/ncee/wwc/help/idocviewer/Doc.aspx?docId=19&tocId=7>.

groups; the differential attrition rate was 6.1 percent. This amount of attrition is within the acceptable range established by WWC standards.

Weaknesses:

- None.

Summary of 2009 Striving Readers Projects: Profile of Michigan Department of Education Striving Readers Project and Evaluation¹

Grantee: Michigan Department of Education

Project Director: Ruth Isaia

Local Evaluator: SRI International

Principal Investigators: Jose Blackorby, Ellen Schiller

Project Website: <http://policyweb.sri.com/cehs/projects/displayProject.jsp?Nick=mdesrp>

Setting of the Study

The seven middle and high schools participating in the Michigan Striving Readers program were located in three districts in the southeast and western suburban areas of Michigan: Muskegon Public Schools, Inkster Public Schools, and Westwood Community School District. There were four middle schools and three high schools in the study from these districts. In the 2008-2009 school year, participating schools in these districts served 3,824 middle and high school students. The percentage of students eligible for free/reduced price lunches ranged from 51 to 96 percent. The percentage of students reading below proficiency on the 2009 Michigan Educational Assessment Program (MEAP) reading test ranged from 26 to 61 percent, with an average of 42 percent across the seven schools. The Striving Readers project was funded for two years, including one year of implementation of the intervention during the 2010-11 school year.

Intervention Model

Classroom Model as Planned²

The Fusion Reading Program was developed by researchers at the University of Kansas Center for Research on Learning (KU-CRL).³ Fusion is a two-year, intensive supplemental reading course designed for middle and high school struggling readers who score at least two years below grade level on standardized reading measures. Struggling students are enrolled in the intervention for one class period for five days a week. The intervention is a highly-structured course designed to teach an array of reading strategies within a scaffolded scope and sequence of instruction, practice, feedback, and ongoing assessments for progress monitoring.

¹ See final report: Schiller, E., Wei, X., Thayer, S., Blackorby, J., Javitz, H., & Williamson, C. (2012). *A Randomized Controlled Trial of the Impact of the Fusion Reading Intervention on Reading Achievement and Motivation for Adolescent Struggling Readers*. Arlington, VA: SRI International.

² For more information on *Fusion*, please see the Fusion Reading Program Intervention Profile by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

³ Fusion builds on the work of the KU Strategic Instruction Model's Learning Strategies Curriculum and Xtreme Reading by integrating some of the same strategies, focusing on reading, and extending the time frame from 1 to 2 years in duration.

Reading instruction in the Fusion curriculum is built around nine units and a student project: (1) the *Establish the Course* unit provides students with rationales for the course, an overview of course content, and expectations for classroom management; (2) the *Prediction Strategy* unit explicitly teaches students the first reading comprehension strategy, and students learn how to preview reading selections, link prior knowledge to the subject, make predictions and inferences about content, and evaluate reading to answer student generated questions and predictions; (3) the *Possible Selves* unit surfaces long-term future goals and establishes action plans that link attainment of personal goals to reading proficiency; (4) the *Bridging Strategy* unit provides instruction in advanced phonics, decoding, word recognition, and reading fluency; (5) the *Strategy Integration* unit teaches students how to integrate prediction, bridging, and vocabulary strategies and provides students with opportunities to apply integrated strategies to reading content area textbooks; (6) the *Summarization Strategy* unit teaches students to summarize small sections of books, chapters, and some longer passages; (7) the *Strategy Integration* unit continues teaching and providing opportunities for students to practice integrating strategies and applying them to reading; (8) the *PASS the Test* unit teaches students a reading strategy they can use to do well on standardized tests; (9) the *Advanced Strategy Integration* unit continues teaching and providing opportunities for students to practice integrating strategies and applying them to reading; and (10) students do a final Fusion Reading intervention project to apply the reading strategies they have learned.

Instructional routines follow a specific sequence: (1) Warm-Ups, where students are engaged in an activity at the beginning of class to provide a connection to class readings and key strategies; (2) Thinking Reading, a structured process in which the teacher demonstrates reading behaviors so that students can participate in the process; (3) Explicit Instruction, where, for each strategy, teachers describe, explain, and model specific metacognitive steps of the strategy and students (a) verbally practice the steps of the strategy and practice using the strategy first with materials at their instructional level and later with increasingly difficult materials and (b) receive elaborated feedback from the teacher until they gain proficiency and are able to use the strategy in a generative way and apply the strategy to assignments in a wide variety of materials and settings; (4) Explicit Vocabulary Instruction, which follows a seven-step vocabulary process; and (5) Lesson Review.

Progress assessment forms and answer sheets are provided at the beginning and end of each Fusion Book, and formative assessment activities are available during partner and individual practice sessions throughout each unit.

Professional Development Model as Planned

The professional development model for teachers in the Fusion program included initial whole group training, a mid-year workshop, a follow-up workshop toward the end of the school year, and at least 40 hours of in-class support. In total, nine Fusion intervention teachers (one per school but two in each of the two large high schools) were required to attend seven days of group professional development training. All training was administered by professionals certified by the International Professional Development Network (IPDN), an organization created by KU-CRL.

In August 2010, Fusion teachers received an initial three days of professional development. This training included instruction on classroom routines, classroom set-up, and instructional methodology. During the fall semester, professional developers from IPDN were available to work with intervention teachers to respond to questions, monitor intervention progress, and provide on-going support. In November or December of the implementation year, a two-day workshop for teachers was provided, covering follow-up and problem solving, an introduction to the motivation component of the intervention, site visits, and feedback. A second two-day workshop took place in late winter/early spring. This session continued training on follow-up and problem solving (begun in the previous session) and provided training in the final instructional components of the Fusion intervention.

Once formal instruction began, 40 hours of instructional coaching was provided to each intervention teacher according to the Instructional Coaching model developed by KU-CRL. Throughout the year, teachers videotaped their instruction. Using an online tool, teachers and the professional developers viewed teachers' delivery of the intervention simultaneously, and the professional developers had the opportunity to provide additional feedback to teachers. Each teacher participated in two of these sessions during the school year.

Context for Implementation

Approximately 750 struggling 6th-10th grade readers participated in the Michigan Striving Readers study in the first year, with half of the students randomly assigned to receive the intervention for a two-year period. Random assignment occurred in the spring before the intervention year. For those students assigned to the treatment group, the Fusion intervention replaced an elective class and was delivered as a daily 45 or 90 minute supplement to regular English Language Arts classes.

Students eligible to participate in the study were identified using a two-step process. In the spring preceding each implementation year, the Michigan Educational Assessment Program (MEAP) scores of students entering middle and high school were reviewed. Students scoring in the lowest two categories - "Approaching Proficient" and "Non-Proficient" - on either the narrative or informational selection of the MEAP reading test were administered the Passage Comprehension subtest of the Test of Silent Contextual Reading Fluency (TOSCRF) by KU-CRL staff. Students who scored below proficient on the MEAP and/or placed between the 5th and 35th percentile on the TOSCRF were eligible for the study. In addition, students new to the schools in the fall were screened with the TOSCRF to determine eligibility and were randomly assigned to intervention or control within 10 days of the start of school.

Special education students who were classified as cognitively impaired, who had severe multiple disabilities, or who had sensory impairments/autism were excluded from participating in the study. In addition, English Language Learners (ELL) classified as Basic Level 1A were not eligible for the study.

Evaluation Design

Research Questions

1. What is the impact of Fusion on the reading outcomes of struggling readers after receipt of one year of the two-year intervention?
2. What is the fidelity of implementation of Fusion in participating schools?

Impact Study

Research Design and Methods: The impact of the first year of the Fusion intervention was evaluated using an experimental design and within-school randomization of students. Analyses were conducted using a multi-level HLM design with treatment, schools, and grade modeled as fixed effects. Analyses were conducted separately for middle and high school students as well as for all students combined.

Analyses were conducted for each of six outcome domains: word attack, oral reading fluency, reading comprehension, vocabulary, general literacy achievement, and reading motivation. Pretest scores on each measure were included in the models. Additional covariates included race/ethnicity, free and reduced price lunch status, and gender.

Control Condition: Students assigned to the control condition had the opportunity to participate in non-literacy based electives or in study hall. None of the schools in the study had any existing supplemental literacy programs that were available to students in the control condition. Each school's schedule of classes and extracurricular activities were examined to ensure that no other classes with explicit literacy instruction were added to the course offerings.

Sample Size: The sample that was randomly assigned included 366 treatment and 388 control students in grades 6-10. For the Test of Word Reading Efficiency (TOWRE) Phonetic Decoding Efficiency and Sight Word Efficiency scores, the sample for the impact analysis of one year of the intervention included 283 treatment students and 297 control students in grades 6-10. For the Group Reading and Diagnostic Evaluation (GRADE) Passage Comprehension and Sentence Comprehension scores, the sample for the impact analysis of one year of the intervention included 286 treatment students and 296 control students in grades 6-10. For the GRADE Vocabulary score, the sample for the impact analysis of one year of the intervention included 287 treatment students and 296 control students in grades 6-10. For the Michigan Educational Assessment Program (MEAP) score, the sample for the impact analysis of one year of the intervention included 118 treatment students and 138 control students in grades 6-10. For the reading motivation score, the sample for the impact analysis of one year of the intervention included 273 treatment students and 283 control students in grades 6-10.

The sample of students for the impact analysis is large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .12 on the TOWRE Phonetic Decoding Efficiency subtest score after one year of the intervention for grades 6-10,⁴
- .13 on the TOWRE Sight Word Efficiency subtest score after one year of the intervention for grades 6-10,
- .21 on the GRADE Sentence Comprehension subtest score after one year of the intervention for grades 6-10,
- .21 on the GRADE Passage Comprehension subtest score after one year of the intervention for grades 6-10,
- .18 on the GRADE Vocabulary subtest score after one year of the intervention for grades 6-10,
- .26 on the MEAP score after one year of the intervention for grades 6-10, and
- .20 on the Children’s Academic Intrinsic Motivation Inventory after one year of the intervention for grades 6-10.

Key Measures of Student Reading Outcomes (Source)

Test of Word Reading Efficiency (TOWRE): Phonetic Decoding Efficiency and Sight Word Efficiency subtests (External Test Publisher)

Group Reading and Diagnostic Evaluation (GRADE): Sentence Comprehension, Passage Comprehension, and Vocabulary subtests (External Test Publisher)

Michigan Educational Assessment Program (MEAP): Reading (State Test)

Children’s Academic Intrinsic Motivation Inventory (CAIMI) (External Test Publisher)

Implementation Study

Research Design and Methods: The fidelity of implementation of the intervention model was studied in the seven treatment schools. Fidelity of implementation of the professional development model for each intervention teacher was not calculated.

The fidelity of implementation of the classroom model was based on the number of Fusion lessons teachers covered in two semesters or two trimesters of teaching the curriculum (both of which contain approximately the same amount of allocated classroom time). The proportion of curriculum coverage was calculated as the number of strategies covered by the teacher relative to the total number of lessons teacher should have covered during the first year of Fusion. Adequate implementation was defined as covering at least 80 percent of the lessons.

Fusion developers suggested that to maximize the full potential of the intervention, the classes needed to be scheduled daily, students should attend at least 80 percent of the allocated class time and class size, and schools should, on average, enroll between 15 and 18 students per class. In the implementation in Michigan, the Fusion teachers entered the actual time students attended classes each day. A student’s dosage rate was calculated as the proportion of time a student was present in class relative to the total number of allocated class minutes. Adequate dosage was defined as attending at least 80 percent of the Fusion classes.

⁴ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

Evaluation Findings

Fidelity of Implementation of the Intervention Model

For the classroom model, the average proportion of lessons covered was 73 percent. Only a third of the teachers covered at least 80 percent of the lessons they were expected to cover in the first year. For student dosage, the average rate of attendance in the Fusion classes was 73 percent. 57 percent of students attended at least 80 percent of the Fusion classes.

Impact of the Intervention on Student Reading Outcomes

There were statistically significant impacts on two measures of reading achievement of struggling readers in grades 6-10 after one year of implementation of the two-year Fusion intervention, with effect sizes of .10 for the Sight Word Efficiency score on the TOWRE and .15 for the Sentence Comprehension score on the GRADE. There were no statistically significant impacts on the other scores. The effect sizes were .005 on the Phonetic Decoding Efficiency score of the TOWRE, .01 on the Passage Comprehension score of the GRADE, -.009 on the Vocabulary score on the GRADE, and .11 on the MEAP reading score. There was no statistically significant impact on student motivation to read. The effect size on the CAIMI was .01.

Readers should exercise caution when drawing conclusions based on the lack of statistical significance of the findings. Because funding for the Striving Readers program was cancelled after one year of data collection, the study did not have an adequate sample size (i.e., number of students) to reliably detect policy-relevant impact of Fusion on one of the reading outcomes of interest (MEAP). The originally planned study design described in the grant application would have had a much larger sample size based on three years of data collection.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Intervention

Strengths:

- The analysis of the impact of one year of Fusion on reading achievement (TOWRE, GRADE, and MEAP) meets What Works Clearinghouse (WWC) evidence standards. Analyses that meet WWC evidence standards make us the most confident that the effect we see is due solely to the intervention alone, and not to the many other factors that are at play in schools and in the lives of students, such as teachers, school and family.
- Eligibility for random assignment was determined systematically, using a predetermined cutoff score on two tests of reading achievement (the Michigan Educational Assessment Program (MEAP) and the Test of Silent Contextual Reading Fluency (TOSCRF)).
- Random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.

- There was no evidence that other factors (e.g., other reading programs or district policies) were implemented in ways that would have undermined the evaluators' ability to attribute impacts to the Fusion Reading Program.
- The evaluation employed three reading tests as outcome measures. The Group Reading Assessment and Diagnostic Evaluation (GRADE) Sentence Comprehension, Passage Comprehension, and Vocabulary subtests; the Test of Word Reading Efficiency (TOWRE): Phonetic Decoding Efficiency and Sight Word Efficiency subtests; and the Michigan Educational Assessment Program (MEAP) Reading test. There was no reason to believe that students assigned to the treatment group had more experience taking the tests than did the control group students, or that the tests measured skills specific to the intervention, both of which could have undermined confidence in the impact estimates.
- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement was included in the models to increase the precision of the impact estimates.
- While some students were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggested that the integrity of the original randomized design was preserved, and that treatment and control groups continued to be statistically equivalent on all measured and unmeasured characteristics at follow-up.
 - 23.2 percent of grade 6-10 students who received one year of Fusion were unable to participate in follow-up data collection on the TOWRE subtests. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.9 percent. This amount of attrition is within the acceptable range established by WWC standards.⁵
 - 22.9 percent of grade 6-10 students who received one year of Fusion were unable to participate in follow-up data collection on the GRADE Passage Comprehension and Sentence Comprehension subtests. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.6 percent. This amount of attrition is within the acceptable range established by WWC standards.
 - 22.8 percent of grade 6-10 students who received one year of Fusion were unable to participate in follow-up data collection on the GRADE Vocabulary subtest. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.2 percent. This amount of attrition is within the acceptable range established by WWC standards.

⁵ For more information, please see Appendix A-Assessing Attrition Bias, of the WWC Procedures and Standards Handbook, available at: <http://ies.ed.gov/ncee/wwc/help/idocviewer/Doc.aspx?docId=19&tocId=7>.

Weaknesses:

- Abt was unable to calculate student-level attrition for the sub-sample of students tested using the MEAP. The evaluator did report data that established that the analytic sample of student used to estimate the impact of one year of Fusion on MEAP scores were equivalent on the MEAP at baseline.

Summary of 2009 Striving Readers Projects: Profile of New York Department of Education Striving Readers Project and Evaluation¹

Grantee: New York State Department of Education

Project Director: Colleen Ferrone

Local Evaluator: State University of New York at Albany

Principal Investigator: Dianna Newman

Project Website: <http://www.albany.edu/education/2683.php>

Setting of the Study

Thirteen New York City middle schools participated in the New York Striving Readers program. All of the middle schools were Title 1-eligible and had a significant percentage of students who were English Language Learners (ELL); seven of the schools had an ELL student population that exceeded 20 percent. The Striving Readers project was funded for two years, which included one year of implementation of the intervention during the 2010-11 school year.

Intervention Model

Classroom Model as Planned²

The New York Striving Readers program implemented the REWARDS Program, developed by Dr. Anita Archer and Dr. Mary Gleason. The REWARDS Program provides comprehensive instruction in word analysis, fluency, vocabulary, reading comprehension and writing; and uses content-related text and extended discussion of text meaning and interpretation to enhance student motivation and engagement in literacy learning. The REWARDS Program includes: (1) REWARDS-Multisyllabic Word Reading Strategies; (2) REWARDS Plus (which applies reading strategies to content area passages in science and social studies in order to enhance content knowledge while expanding reading strategies); (3) REWARDS WRITING-Sentence Refinement (which builds writing skills); and (4) REWARDS Plus (additional Application Lessons based on social studies or science passages). The three components in the REWARDS Program are taught in an integrated sequence. Students begin with REWARDS and complete all lessons before moving on to REWARDS Plus Social Studies and then REWARDS Plus Science. Students receive REWARDS Writing throughout the year.

¹ No final evaluation report was completed because funding ended after one year. No impact report was produced, but an interim implementation report was available: Kundert, D.K. (2012). *Striving Readers Project. New York State Department of Education/New York City Department of Education. Fidelity of Classroom Implementation Report and Implementation Fidelity Score Report 2010-11*. Albany, NY: Evaluation Consortium, University at Albany, State University of New York.

² For more information on *REWARDS Suite*, please see the REWARDS Suite Intervention Profile by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

The REWARDS Program was used for eligible seventh-graders in New York City middle schools in a supplemental daily period of intervention for one school year. Students assigned to the intervention meet for 40-50 minutes per day.

Professional Development Model as Planned

Developers of the REWARDS intervention provided formal staff development, on-site coaching, and follow-up training sessions over the course of the implementation year. Altogether, teachers were provided over 60 hours of formal training. Teachers were required to participate in a five-day Summer Institute prior to the first implementation year. During this initial group training, interventionists were introduced to essential instructional practices and the structure and content of the REWARDS Program. One half-day of the Summer Institute was dedicated to training interventionists how to use the online formative assessments. To provide a forum for group feedback and follow-up training, two Saturday training sessions were held during the course of the academic year. There was also a series of school-based meetings and five informational meetings teachers were required to attend.

In addition to the initial group training, interventionists received a minimum of two visits from the developers of the program, to offer in-class support and coaching. Each teacher also received three on-site coaching visits by a project coach. School principals and other district staff also visited each intervention classroom at least twenty times a year to ensure that implementation was progressing appropriately, and to provide any necessary support to teachers.

Context for Implementation

Approximately 526 grade 7 students in the selected middle schools participated in the study each year, with 257 students randomly assigned to receive the intervention and 269 assigned to the control group. Eligibility for the study was determined by a two-phase approach. First, in the spring of 2010, 6th grade students who scored at Level 2 on the Grade 5 NY English Language Arts Assessment were given the Gates-MacGinitie Reading Tests (GMRT). Those students scoring between 1 and 2.5 standard deviations below the mean for that age group were provisionally eligible. Then, in the fall of 2010, when Grade 6 NY ELA scores became available, eligibility was reassessed. Students scoring above a Level 2 were no longer eligible, regardless of their performance on the GMRT the previous spring. Students who scored at a Level 2 on the Grade 6 NY ELA were administered the GMRT and were eligible for the study if they scored between 1 and 2.5 standard deviations below the mean for that age group. English Language Learners who did not pass the New York State English as a Second Language Achievement Test (NYSESLAT) and Special Education students whose IEPs precluded them from participating in the intervention were excluded from the study.

Evaluation Design

Research Questions

1. Does participation in the REWARDS Program improve reading achievement for struggling readers in grade 7 with one year of exposure to the intervention?

2. What is the fidelity of implementation of the REWARDS program in the study classrooms?

Impact Study

Research Design and Methods: The impact of the *REWARDS* intervention on student reading achievement was determined using an experimental research design and within-school random assignment of students.

Control Condition: Students assigned to the control condition attended regular ELA classes along with their peers in the treatment group and classmates not participating in the study. While students in the treatment group received REWARDS in addition to their regular ELA classes, control group students were assigned to a non-ELA elective class as part of their regular 7th grade schedule. The evaluator reviewed the overall student course of study for all control students to make sure that no other ELA course or special instruction in ELA - beyond what is in the regular ELA classroom - occurred.

Sample Size: The samples that were randomly assigned included a treatment group of 257 struggling readers in 7th grade who received REWARDS for one year, and 269 struggling readers in the control group, across 13 New York City middle schools. On the Gates-MacGinitie Reading Tests (GMRT) Total Reading and Comprehension subtest, the sample for the impact analysis of one year of the intervention included 228 treatment students and 235 control students in 7th grade. On the GMRT Vocabulary subtest, the sample for the impact analysis of one year of the intervention included 232 treatment students and 237 control students in 7th grade. On the New York State English Language Arts Assessment, the sample for the impact analysis of one year of the intervention included 253 treatment students and 264 control students in 7th grade.

The sample of students for the impact analysis is large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .21 on the New York State English Language Arts Assessment after one year of the intervention for grade 7,³
- .19 on the GMRT Total Reading test after one year of the intervention for grade 7,
- .22 on the GMRT Comprehension subtest after one year of the intervention for grade 7, and
- .19 on the GMRT Vocabulary subtest after one year of the intervention for grade 7.

Key Measures of Student Reading Outcomes (Source):

New York State Grade 7 English Language Arts Assessments (State Test)

Gates-MacGinitie Reading Tests (GMRT): Vocabulary, Comprehension, and Total Reading (External Test Publisher)

³ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

Implementation Study

Research Design and Methods: The fidelity of implementation of the intervention model was studied in the 13 treatment schools. To determine fidelity of the professional development model, two scores were computed: one based on the professional development observation and a second based on the teacher's level of participation in the professional development activities. The professional development fidelity observation assessed the fidelity of the professional development training provided to teachers and administrators, across nine content areas. The evidence of training for each area was scored on a four-point scale (none, low, moderate, or high evidence), and a total subscore for professional development was calculated by computing an average score. This average score was the basis for a four-level fidelity rating: extremely limited fidelity (0-0.74), limited fidelity (0.75-1.49), moderate fidelity (1.50-2.24), or high fidelity (2.25-3.00). An attendance score was calculated for each teacher, using the number of days attended divided by the total number of days of professional development offered. A total participation score was calculated based on the percentage attendance across all of the professional development activities. Four levels of attendance were defined: high attendance (75-100 percent), moderate attendance (50-64 percent), limited attendance (25-49 percent), and extremely limited attendance (less than 24 percent). An overall score for fidelity of professional development was calculated by summing the total scores for level of participation and implementation of professional development, creating a 6-point scale. Level of overall fidelity was defined as follows: extremely limited fidelity (0-1.49), limited fidelity (1.5-2.99), moderate fidelity (3.00-4.49), and high fidelity (4.5-6.0).

For fidelity of the classroom model, moderate attendance was based on classroom observation and level of student attendance. For the observation, five domains of instruction were rated on their implementation as high, moderate, low, or no evidence. A similar four-level score was used to rate the degree of evidence for teacher and student behaviors. These two scores were averaged to create the classroom fidelity score: high fidelity (2.25-3), moderate fidelity (1.5-2.24), limited fidelity (0.75-1.49), or extremely limited fidelity (0-0.74). For student attendance, each student was scored on their percentage of days attended and an average was calculated for the classroom. Fidelity of student attendance was scored as follows: high fidelity (75-100 percent average attendance), moderate fidelity (50-75 percent average attendance), limited fidelity (25-49 percent average attendance), and extremely limited fidelity (0-24 percent average attendance).

Evaluation Findings

Fidelity of Implementation of the Intervention Model

For fidelity of professional development, the overall score for attendance for the intervention teachers was 2.91 and the overall score for implementation of professional development was 2.86. Together, the total score was 5.77, which corresponded to a high level of fidelity for the professional development model.

For the classroom model, based on observations of the instruction, the level of classroom implementation corresponded to a high fidelity level (score = 2.78 out of 3). Overall, the REWARDS program was delivered consistently by teachers with high fidelity (score = 2.87 out of 3). In terms of student behavior, students demonstrated effective learner behaviors across most

of the observations. The overall score was 2.83 out of 3, which corresponded to a high level of fidelity. For student attendance, the average percentage of days in attendance was 78.4 percent, which corresponded to a high level of fidelity of attendance.

Impact of the Intervention on Student Reading Outcomes

There was not a statistically significant impact on the reading achievement of struggling 7th grade students with one year of exposure to the REWARDS program. On the New York state English Language Arts test, the effect size was .15. The effect sizes on the GMRT Vocabulary, Comprehension, and Total Reading scores were .08, -.01, and .02 respectively.

Readers should exercise caution when drawing conclusions based on the lack of statistical significance of the findings. Because funding for the Striving Readers program was cancelled after one year of data collection, the study did not have an adequate sample size (i.e., number of students) to reliably detect some policy-relevant impacts of REWARDS on the New York state English Language Arts test. The originally planned study design described in the grant application would have had a much larger sample size based on three years of data collection.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Intervention

Strengths:

- The analysis of the impact of one year of REWARDS on reading achievement (NYSELAA and GMRT) meets What Works Clearinghouse (WWC) evidence standards. Analyses that meet WWC evidence standards make us the most confident that the effect we see is due solely to the intervention alone, and not to the many other factors that are at play in schools and in the lives of students, such as teachers, school and family.
- Eligibility for random assignment was determined systematically, using dual predetermined cutoff scores on tests of reading achievement (the Gates-MacGinitie Reading Test and the New York state English Language Arts Assessment).
- Random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.
- There was no evidence that other factors (e.g., other reading programs or district policies) were implemented in ways that would have undermined the evaluators' ability to attribute impacts to the REWARDS Program.
- The evaluation employed two reading tests as outcome measures. The Group Gates-MacGinitie Reading Test (GMRT) Total Reading, Comprehension, and Vocabulary tests and the New York State English Language Arts Assessment. There was no reason to believe that students assigned to the treatment group had more experience taking the tests than did the control group students, or that the tests measured skills specific to the intervention, both of which could have undermined confidence in the impact estimates.

- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement was included in the models to increase the precision of the impact estimates.
- While some students were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggested that the integrity of the original randomized design was preserved, and that treatment and control groups continued to be statistically equivalent on all measured and unmeasured characteristics at follow-up.
 - 12.0 percent of grade 7 students who received one year of the REWARDS Program were unable to participate in follow-up data collection on the GRMT Total Reading and Comprehension tests. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.4 percent. This amount of attrition is within the acceptable range established by WWC standards.⁴
 - 10.8 percent of grade 7 students who received one year of the REWARDS Program were unable to participate in follow-up data collection on the GRMT Vocabulary test. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 2.1 percent. This amount of attrition is within the acceptable range established by WWC standards.
 - 1.7 percent of grade 6 students who received one year of the REWARDS Program were unable to participate in follow-up data collection on the New York State English Language Arts Assessment. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 0.3 percent. This amount of attrition is within the acceptable range established by WWC standards.

Weaknesses:

- None.

⁴ For more information, please see Appendix A-Assessing Attrition Bias, of the WWC Procedures and Standards Handbook, available at: <http://ies.ed.gov/ncee/wwc/help/idocviewer/Doc.aspx?docId=19&tocId=7>.

Summary of 2009 Striving Readers Projects: Profile of Virginia Department of Education Striving Readers Project and Evaluation¹

Grantee: Virginia Department of Education
Project Director: Kendall Hunt
Local Evaluator: RMC Research Corporation
Principal Investigator: Allen Schenck
Project Website: <http://www.doe.virginia.gov>

Setting of the Study

Nine schools across three school districts were selected to participate in the Virginia Striving Readers Intervention Initiative (VSR II). The school districts participating in the project included Norfolk City Public Schools, Richmond City Public Schools, and Roanoke City Public Schools. In the 2008-2009 academic year, the nine schools in these three districts served 5,492 students in grades 6-8. Students eligible for free and reduced price meals comprised 61 percent of the school population. Six of the schools participating in the VSR II program did not make Adequate Yearly Progress (AYP) in the 2008-2009 academic year. The Striving Readers project was funded for two years, including one year of implementation of the intervention during the 2010-11 school year.

Intervention Model

Classroom Model as Planned²

Voyager Passport Reading Journeys (PRJ) is a Cambium Learning Group curriculum. The program is organized into daily, 50-minute lessons with explicit, systematic instruction in critical language skills. It is formatted as a series of lessons, or Expeditions, designed to be delivered over the course of one school year on topics related to science, social studies, and literature. PRJ also incorporates video segments on DVD, text, and online interactive lessons. PRJ contains different levels of instruction for different grades. Each level is designed for one full academic year with 15 two-week Expeditions with optional add-ons (reteach or writing). While each level uses similar strategies and reinforces common skills, there are unique Expedition topics and content that is age/grade appropriate and of increasing complexity for higher grades. The intervention is called PRJ I for seventh graders and PRJ II for eighth graders.

¹ See final report: Schenck, A., Jurich, S., Frye, M., Lammert, J., & Sayko, S. (2012). *Evaluation Report/Impact Study: Virginia Striving Readers Intervention Initiative (VSR II)*. Arlington, VA: RMC Research.

² For more information on PRJ, please see the Voyager Passport Reading Journeys Intervention Profile by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

The Expeditions focus on six instructional practices: (1) explicit vocabulary instruction; (2) direct and explicit comprehension strategy instruction; (3) extended discussion of text meaning and interpretation; (4) instruction in reading foundational skills, such as decoding and fluency; (5) instruction in writing; and (6) strategies to improve student motivation and engagement in literacy learning. PRJ blends reading foundational skills, vocabulary instruction, direct and explicit comprehension strategies, text meaning and interpretation, and writing. The intervention is based on reading research and research in learning. Instruction in reading foundational skills is provided through the word study component of the Expeditions. The students with the lowest word reading abilities are taught with a thirty-lesson word study program, beginning with a review of single letter-sound correspondences. These explicit word study lessons may either be provided prior to implementing the first Expedition lesson or on alternate days once the intervention sequence has begun. Explicit instruction is delivered with a focus on decoding, spelling, vocabulary, comprehension, and fluency. In addition, the students receive continued systematic and explicit instruction in practices that teach them to be flexible decoders. Lessons offer instruction in affixes, sight words, decoding multisyllabic words, spelling, and word or phrase fluency.

Vocabulary instruction is addressed through the use of explicit instruction of word meanings and development of strategies to determine unknown words through morpheme analysis. A planned sequence of vocabulary skills and multiple exposures to high-utility words are meshed within the passages, comprehension activities, and text discussions. Affixes and roots are explicitly taught to students in a sequential pattern that is supported by the identified words in the passages. *SOLO* is a computer program that provides self-paced practice on vocabulary and comprehension skills. There are multiple tools that help students determine word meaning and contextual use in self-selected, Lexile-leveled³ reading passages. New words are introduced with age-appropriate definitions and examples. Supports include automated clues or prompts and a function that allows students to click on difficult words to hear their pronunciation and definition.

Direct and explicit comprehension strategies are woven into instruction to help students develop skills that are traditionally lacking among striving readers, such as making and confirming predictions, identifying or generating main ideas, summarizing, and making inferences. Comprehension skills are taught explicitly and applied to expository passages both in the text and in *SOLO*. The lessons incorporate strategies for making connections, asking questions, visualizing, and making inferences. Students also examine organizational text features that serve as frames for information and logical links between ideas. Comprehension strategies are scaffolded in three stages: interventionist modeling, interventionist assistance with student practice, and student independence. These stages represent a gradual shift in the responsibility for learning from the interventionist to the students. Direct instruction includes modeling, in which the interventionist reads aloud to show students how to use the reading strategies. Interventionists think aloud with students, to make thought processes transparent to students. Modeling is followed with direct, guided practice and self-assessment that enables students to apply the newly learned skills and strategies in a variety of texts and reading levels.

³ Lexile is a numeric representation of an individual's reading ability or a text's readability based on the work of Jackson Stenner, from MetaMetrics, Inc.

Discussion of the *text meaning and interpretation* are elicited through questions posed by the interventionist during and after reading. In the first reading of the selection, the interventionist asks literal comprehension questions to ensure understanding and model the metacognitive process of self-monitoring. After students complete their reading, the interventionist asks critical thinking questions. This interventionist-directed questioning is integrated with student-generated questions, as a key reading comprehension strategy during reading and a way for students to monitor and deepen their understanding of the text.

A library of age-appropriate Lexile-leveled books and magazines is provided for each intervention classroom. The primarily nonfiction texts selected have been tested for high interest with middle and high school students. Lessons include whole-group instruction as well as flexible grouping to ensure maximum use of differentiated instructional time.

Professional Development Model as Planned

For the Virginia Striving Readers study, instruction was provided by trained interventionists with support from Voyager Implementation Specialists. The professional development model for the eight interventionists in the Voyager program (two schools shared one interventionist) consisted of an initial whole group training (16 hours), online self-paced training (two 15-hour online modules on adolescent literacy research that focuses on vocabulary and comprehension for middle school students), additional group training on assessment and other topics (nine hours), and a statewide meeting at the end of the school year (four hours). Implementation specialists from Voyager delivered the training and provided ongoing support for the interventionists. These professional development modules were mandatory for interventionists. Altogether, over the year of implementation, each of the eight interventionists was intended to receive 59 hours of professional development

The school district liaisons and project coordinator planned to conduct unscheduled visits to the PRJ classrooms to monitor implementation and ensure fidelity to the original design of the model. State department staff planned to visit each school at least eight times during the school year to provide technical assistance during and in between Voyager implementation specialist visits to ensure program fidelity. District coordinators planned to provide additional support by coordinating project efforts at the local level and serving as liaisons between and among partners. For the Striving Readers study, at each school, the principal was responsible for ensuring that the interventionists attended all the required training. Principals also planned to visit each intervention teacher once a week during the initial month of implementation and then twice a month over the remaining school year. In addition, Voyager specialists planned to monitor and record the participation of the interventionists and provide feedback on their progress throughout the year. District and LDOE staff would receive a total of 30 and 18 hours of professional development, respectively; principals would receive 36 hours across the first school year of implementation.

In addition, interventions were intended to receive on-site monthly coaching to address the individual needs of the interventionists. The Voyager specialists worked with the designated teachers for a full day, ten days a year. The number of days they were expected to provide coaching was negotiated beforehand with the school districts. Attendance at the year-end

statewide meeting also was required for all interventionists. The expectation was that each interventionist could receive a maximum of 10 days (50 hours) of coaching over the school year.

Context for Implementation

The Virginia Striving Readers program was a one-year intervention that served approximately 900 grade 7 and grade 8 students, with half of the students randomly assigned to receive the PRJ intervention and half assigned to the control group. Random assignment took place in the spring of 2010 at each grade level. For those students assigned to the treatment group, PRJ was provided as a daily, 50 minute supplement to regular English Language Arts classes, replacing an elective class (which the control group students attended at the same time that the treatment students were in the PRJ intervention classes).

Student eligibility was based on student scores on the Gates-MacGinitie Reading Tests (GMRT), which the study schools administered to all students in grades 6 and 7. Students whose reading scores were equivalent to two or more grades below the expected grade level were eligible for the Virginia Striving Readers program. In addition, students in grades 6 and 7 who did not pass the 2010 Virginia Standards of Learning (SOL) English/Reading Assessment in the spring were also eligible for the intervention, regardless of their score on the Gates-MacGinitie Reading Tests.

Students with IEPs who required more intensive, one-on-one instruction and those who required highly specialized interventions (i.e., students with intensive cognitive and/or emotional disabilities who could not benefit from group instruction) were excluded from the study. English Language Learners (ELLs) with very limited English skills (Levels 1 and 2) who required basic English language instruction will were excluded from the study.

Evaluation Design

Research Questions

1. What is the impact of one year of exposure to the PRJ supplemental literacy intervention on the reading achievement of striving readers in grades 7 and 8?
2. What is the fidelity of implementation of the PRJ supplemental literacy intervention in the participating schools?

Impact Study

Research Design and Methods: The impact of the intervention was evaluated at the end of one year of implementation in the 2010-2011 school year, using an experimental design and within-school random assignment of students. Impacts were estimated for grades 7 and 8 combined and separately.

The analytical model that was used to determine the impact of the PRJ intervention was a two-level regression model including student- and school-level variables. The student-level model estimated impacts on end-of-year student reading achievement scores on the GMRT (April) and the SOL English/reading assessment (May) as a function of the school mean and several covariates: whether the student was in the treatment or control group, the student's baseline

reading achievement score in the previous school year on the same test, and other student demographic variables (e.g., gender, race/ethnicity, free/reduced lunch, etc.). The school level model estimated the school reading achievement averages as a function of selected school characteristics (e.g., contextual school aggregates of student demographic variables, enrollment size, etc.). While the school average was modeled as a random effect, the regression coefficient for the treatment variable (i.e., the treatment effect) was assumed to be constant across schools and modeled as a fixed effect.

Control Condition: Students randomly assigned to the control group were able to choose an elective from a variety of offerings including art, music, foreign languages, and enrichment classes. They attended their chosen elective at the same time that the students in the treatment group received the PRJ supplemental literacy intervention. Enrichment classes attended by control group students could include literature circles as long as they did not provide systematic reading instruction. Remediation classes that focused on the SOL English/Reading assessment were provided outside of the school day to students who failed any part of the test, but these classes did not include systematic reading instruction. All students participating in the study attended the same English Language Arts classes as their 7th and 8th grade classmates. None of the nine participating schools had reading specialists or literacy coaches on their staff; therefore, the conventional practice in the schools was to have the regular English teachers provide instructional support to struggling readers that focused on strategies for comprehension and writing, during the regular English classes.

Sample Size: The samples that were randomly assigned included a treatment group of 460 struggling readers in grades 7-8 who received PRJ for one year, and 458 struggling readers in the control group, across nine schools. For the Gates-MacGinitie Reading Tests-4th Edition (GMRT) Total Reading score, the sample for the impact analysis of one year of the intervention included 279 treatment students and 289 control students in grades 7-8. For the GMRT Comprehension score, the sample for the impact analysis of one year of the intervention included 279 treatment students and 290 control students in grades 7-8. For the GMRT Vocabulary score, the sample for the impact analysis of one year of the intervention included 290 treatment students and 295 control students in grades 7-8. For the Virginia Standards of Learning English/Reading Assessment (SOL), the sample for the impact analysis of one year of the intervention included 343 treatment students and 358 control students in grades 7-8.

The sample of students for the impact analysis is large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .20 on the GMRT Total Reading score after one year of the intervention for grades 7-8,⁴
- .22 on the GMRT Comprehension score after one year of the intervention for grades 7-8,
- .22 on the GMRT Vocabulary score after one year of the intervention for grades 7-8, and
- .19 on the SOL English/Reading Assessment after one year of the intervention for grades 7-8.

Key Measures of Student Reading Outcomes (Source):

Virginia Standards of Learning (SOL) English/Reading Assessment (State Test)

⁴ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

Gates-MacGinitie Reading Tests-4th Edition (GMRT): Vocabulary, Comprehension, and Total Reading (External Test Publisher)

Implementation Study

Research Design and Methods: The fidelity of implementation of the intervention model was studied in the 10 treatment schools. Level of fidelity of implementation of the professional development model for each intervention teacher was calculated as the number of total PD hours attended relative to the 67 required PD hours. The score was computed by school, since each school had one interventionist, and although they were teaching two different grades, the PD was the same. For instance, if an interventionist attended 51 of the 67 required hours of professional development, the interventionist/school score was 0.76. The evaluators established a score of 1.00 or above as representing adequate fidelity to the professional development model. (An interventionist could achieve a score higher than 1 if they attended more training than was required.) Scores below 1.00 were considered inadequate.

To assess the fidelity of implementation of the classroom model, each intervention teacher was observed twice. Four aspects of classroom instruction were rated: amount of instruction, differentiation of instruction, classroom management, and use of assessments. The four factors were weighted and summed to 100 percent. Fidelity of implementation of the classroom model was defined as: low = 0-69 percent; medium = 70-89 percent; high = 90-100 percent.

Evaluation Findings

Fidelity of Implementation of the Intervention Model

During the first year of implementation, each interventionist received between 72 and 127 hours of professional development that included face-to-face and online training, and individual supports from state and program staff. Based on the required group professional development hours, the index of fidelity of professional development ranged from 0.5 to 1.3. All interventionists, except for school 4, received an index at or above 1.0, which was defined as adequate.

For fidelity of implementation of the classroom model, based on the adequacy of delivery of the model observed during the site visits, two of the nine interventionists were classified as attaining high fidelity of implementation (score above 0.9), while the remaining seven attained adequate (medium) fidelity of implementation levels.

Impact of the Intervention on Student Reading Outcomes

There were no statistically significant impacts on the reading achievement of struggling readers in grades 7-8 after one year of exposure to PRJ, with effect sizes of .06 for the Total Reading score on the GMRT, .05 for the Comprehension score on the GMRT, .07 for the Vocabulary score on the GMRT, and .06 on the Virginia SOL English/Reading score.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Intervention

Strengths:

- The analysis of the impact of one year of PRJ on reading achievement (GMRT and Virginia SOL English/Reading Assessment) meets What Works Clearinghouse (WWC) evidence standards. Analyses that meet WWC evidence standards make us the most confident that the effect we see is due solely to the intervention alone, and not to the many other factors that are at play in schools and in the lives of students, such as teachers, school and family.
- Eligibility for random assignment was determined systematically, using a predetermined cutoff score on a test of reading achievement (the Gates-MacGinitie Reading Tests (GMRT)).
- Random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.
- There was no evidence that other factors (e.g., other reading programs or district policies) were implemented in ways that would have undermined the evaluators' ability to attribute impacts to Voyager Passport Reading Journeys (PRJ).
- The evaluation employed two reading tests as outcome measures. The Gates-MacGinitie Reading Tests (GMRT): Total Reading, Comprehension, and Vocabulary subtests, and the Virginia Standards of Learning Reading Assessment, a measure of general literacy. There was no reason to believe that students assigned to the treatment group had more experience taking the test than did the control group students, or that the tests measured skills specific to the intervention, both of which could have undermined confidence in the impact estimates.
- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement was included in the models to increase the precision of the impact estimates.
- While some students were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggested that the integrity of the original randomized design was preserved, and that treatment and control groups continued to be statistically equivalent on all measured and unmeasured characteristics at follow-up.
 - 38.1 percent of grade 7-8 students who received one year of PRJ were unable to participate in follow-up data collection on the GMRT Total Reading score. The level of attrition did not differ substantially across the treatment and control

groups; the differential attrition rate was 2.4. This amount of attrition is within the acceptable range established by WWC standards.⁵

- 38.0 percent of grade 7-8 students who received one year of PRJ were unable to participate in follow-up data collection on the GMRT Comprehension score. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 2.7 percent. This amount of attrition is within the acceptable range established by WWC standards.
- 36.3 percent of grade 7-8 students who received one year of PRJ were unable to participate in follow-up data collection on the GMRT Vocabulary score. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.4 percent. This amount of attrition is within the acceptable range established by WWC standards.
- 23.6 percent of grade 7-8 students who received one year of PRJ were unable to participate in follow-up data collection on the Virginia state Standards of Learning Reading Assessment. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 3.6 percent. This amount of attrition is within the acceptable range established by WWC standards.

Weaknesses:

- None.

⁵ For more information, please see Appendix A-Assessing Attrition Bias, of the WWC Procedures and Standards Handbook, available at: <http://ies.ed.gov/ncee/wwc/help/idocviewer/Doc.aspx?docId=19&tocId=7>.

Summary of 2009 Striving Readers Projects: Profile of Washington Office of the Secretary of Public Instruction Striving Readers Project and Evaluation¹

Grantee: Washington Office of the Superintendent of Public Instruction

Project Director: Cheryl A. Young, Sarah Rich

Local Evaluator: Education Northwest (formerly Northwest Regional Educational Laboratory)

Principal Investigator: Theresa Deussen

Project Website: <http://www.k12.wa.us/Reading/SpecialProjects.aspx>

Setting of the Study

The Washington Striving Readers Program was implemented in five middle/junior high schools in three districts in the western part of the state. The three participating districts were Auburn, Bethel, and Franklin Pierce School Districts. The three middle schools served students in grades 6-8, while the two junior high schools served only 7th and 8th grade students. In 2008-2009, these schools served 3,968 students. All of the schools were eligible for Title I, and between 45 and 74 percent of their students were eligible for free and reduced price lunch. On average, more than half of the schools participating in the Washington Striving Readers program did not meet the state standards in reading in 2008. Four of the five schools serve English Language Learners (ELLs). The Striving Readers project was funded for two years, which included one year of implementation of the intervention during the 2010-11 school year.

Intervention Model

Classroom Model as Planned²

The Washington Striving Readers Program intervention is a comprehensive package that combines two stand-alone interventions: *Phonics Blitz* and *Read to Achieve*. Depending on how eligible students scored on the AIMSweb CBM fluency assessment, they were assigned to two different groups. One group consisted of students who read fewer than 100 words correct per minute and/or were at or below 90 percent accuracy on the AIMSweb assessment; these students received both the *Phonics Blitz* and the *Read to Achieve* interventions.³ Students who read more than 100 words correct per minute and were at least 90 percent accurate were placed in a different group and received only the *Read to Achieve* intervention.⁴ *Phonics Blitz* targets

¹ Deussen, T., Scott, C., Nelsestuen, K., Roccograndi, A., Davis, A. *Washington Striving Readers Year 1 Evaluation Report Draft*. (2012). Portland, OR: Education Northwest.

² For more information on *Phonics Blitz* and *Read to Achieve*, please see the Phonics Blitz Intervention Profile and the Read to Achieve Intervention Profile, both by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

³ In the Washington evaluation report, this group is labeled as Group 1.

⁴ In the Washington evaluation report, this group is labeled as Group 2.

language decoding, while *Read to Achieve* focuses on vocabulary and comprehension. Both programs address fluency and are designed to increase reading achievement.

Washington Striving Readers classrooms were small, with 12 or fewer students for *Read to Achieve* and nine or fewer for *Phonics Blitz*. Each class was taught by a certified teacher trained in both interventions. *Phonics Blitz* consists of 50 lessons, and was taught in one-hour sessions with each lesson lasting 30-45 minutes. The *Read to Achieve* curriculum includes 39 units, with 5 lessons each. The curriculum is divided into two sections, Content Area and Narrative, with the Narrative section encompassing 15 units. Each lesson is designed to take approximately one class period, or 50-65 minutes.

Phonics Blitz uses teacher-led lessons designed such that teachers and students closely follow a lesson plan with activities sequenced in the same order for each lesson: (1) oral reading; (2) phonemic awareness activities; (3) explicit instruction of phonics concepts; and (4) reading words and sentences with a focus on accuracy, which is intended to be achieved through immediate and positive error correction. Each lesson includes explicit instruction in phonemic awareness, decoding and fluency. Students practice segmenting phonemes orally, which is expected to be a skill that becomes the basis for learning to decode. Phonics concepts are intended to be taught in systematic, explicit, structured lessons using multi-sensory techniques. Fluency is practiced by having students read passages aloud in groups of three for one minute at the beginning of each lesson. Readers then track their accuracy and words correct per minute on a tracking chart.

Read to Achieve includes lessons intended to emphasize comprehension, vocabulary, and fluency strategies as well as higher order thinking skills. In each unit, lessons move from activities in which teachers provide strong support to activities with more moderate levels of teacher support, and eventually, activities that are student-directed, either independently or with a partner. The program incorporates student self-assessment, small group collaboration, and group and individual responses. *Read to Achieve* intends to focus on providing direct, explicit comprehension instruction in every lesson. The skills and strategies that are emphasized in the lessons are: (1) text connections; (2) text structures; (3) comprehension monitoring strategies; (4) note-taking; and (5) metacognitive strategies. *Read to Achieve* also dedicates part of each lesson to practicing decoding and reading fluency. There are two modules in the *Read to Achieve* program: “Comprehending Content Area Text,” which uses content examples that reflect the structure and layout of popular science and social studies texts; and “Comprehending Narrative Text,” which uses non trade book fiction and nonfiction trade books.

Professional Development Model as Planned

The Washington Striving Readers professional development component for the six intervention teachers (one in each of four schools and two in one school) began with a four-day, 28 hour group professional development session in August 2010. Days 1 and 2 consisted of training on *Phonics Blitz*, as well as background instruction in phonics. Days 3 and 4 were devoted to training in the use of *Read to Achieve*. During the school year, there were five additional group trainings lasting approximately seven hours each. Training was provided by the developers of each intervention.

In terms of in-class support, each teacher was visited by the *Phonics Blitz* trainer in October and received an estimated six hours of individualized coaching and support. In addition, the Washington Office of the Superintendent of Public Instruction (OSPI) provided a Striving Readers project coach who observed teachers and gave feedback and support on the implementation of the interventions throughout the year. At a minimum, the Striving Readers project coach was expected to visit each school fourteen times across the school year. Each visit was planned to last approximately two and a half hours, and consisted of observations of both *Phonics Blitz* and *Read to Achieve*, followed by a forty-five minute meeting with the teacher. The principals accompanied the coach on one of the first visits to the intervention classrooms and sat in on the session in order to better understand how the intervention was being implemented.

Context for Implementation

In the first year, the Washington Striving Readers program identified 447 eligible struggling readers in grades 6-8. The students who received *Phonics Blitz* participated in the program during the first twelve weeks of the school year; they then moved into the *Read to Achieve-Comprehending Content Area Text* curriculum for the rest of the year. The students who did not receive Phonics Blitz used the *Read to Achieve-Comprehending Content Area Text* and *Read to Achieve-Comprehending Narrative Text* program for the entire year.

Students were eligible to participate in Washington Striving Readers if they either scored below 390 in the prior year on the Washington Assessment of Student Learning (WASL) or, if they didn't have a state reading assessment score, scored at or below the 32nd percentile on the AIMSweb Maze. All students were screened with the AIMSweb CBM fluency assessment, and students who read below 100 words correct per minute and/or with accuracy below 90 percent were assigned to receive *Phonics Blitz* prior to starting the *Read to Achieve* program.

Students who were eligible for special education in reading and had a reading intervention written into their IEP or who were assigned to a special education resource room the entire day were not eligible to participate in the study. ELLs who scored at a Level 1 on the Washington Language Proficiency Test-II were also excluded from participating in the study. Finally, students who did not have sufficiently strong reading skills to benefit from the program, defined as 6th graders who read below 70 wcpm and 7th graders who read below 75 wcpm and/or had accuracy below 88 percent on the AIMSweb CBM fluency assessment, were excluded.

Evaluation Design

Research Questions

1. Does the Washington Striving Readers program, as a comprehensive package that incorporates a placement assessment and two possible interventions, help struggling middle school readers improve in comprehension and decoding compared to students in a control group?
2. Does a combined intervention of *Phonics Blitz* and *Read to Achieve* help struggling middle school readers with low initial fluency improve in comprehension and decoding compared to students in a control group?

3. Does *Read to Achieve* help struggling middle school readers who read at least 100 words correct per minute with at least 90 percent accuracy improve in comprehension and decoding compared to students in a control group?
4. Does the Washington Striving Readers program, as a comprehensive package that incorporates a placement assessment and two possible interventions, help struggling middle school readers improve in their motivation to read compared to students in a control group?
5. What is the fidelity of implementation of *Phonics Blitz* and *Read to Achieve* in the study classrooms?

Impact Study

Research Design and Methods: The impact of the Washington Striving Readers program was tested using an experimental research design and within-school random assignment of students. The impact of the intervention was analyzed using a two-level HLM model, with schools modeled as fixed effects. The same models were run for each of the outcome measures and with the student's prior score on each measure as a covariate in analyses.

The impact of one year of exposure to the combined intervention package was tested for the full sample treatment versus the control group. In addition, separate impacts of (a) the combined *Phonics Blitz/Read to Achieve-Comprehending Content Area Text* intervention for students entering with low fluency; and (b) the *Read to Achieve-Comprehending Content Area Text and Comprehending Narrative Text* intervention for students who did not require extra practice in decoding will be tested.

Control Condition: Students randomly assigned to the control group attended a study hall or elective class while students assigned to the treatment group received the literacy interventions. The elective classes for control students were taught by instructors who had not been trained and who were in no way involved in the intervention. School staff were interviewed to determine what elective options should be made available to students in the control group that did not involve a literacy component. In most cases, these classes included typical middle school electives such as art, music, dance, theater, or shop.

Students in the control group attended regular English Language Arts (ELA) classes along with their classmates in the treatment group and their peers who were not in the study. Interviews were conducted to determine whether ELA classes were differentiated by student reading levels and whether participation in a regular ELA class could in any way resemble participation in the intervention. In addition, these interviews ascertained whether the school offered any kind of during- or after-school reading tutoring program and, if so, whether students in either the control or treatment groups received this tutoring. It was not expected that any of the participating schools offered reading tutoring services to their students.

Sample Size: The randomly assigned sample included 447 struggling readers in grades 6-8 from five schools in the Washington were eligible to participate in the study in the first year of implementation of the intervention: 222 students were randomly assigned to the treatment group and 225 students were randomly assigned to the control group. On the Gates-MacGinitie Reading Tests (GMRT), the sample for the impact analysis of one year of exposure to the

combined interventions included 176 treatment and 182 control students in grades 6-8. On the Woodcock Reading Mastery scores, the impact analysis sample included 175 treatment and 182 control students in grades 6-8 who received one year of exposure to the combined interventions. The analysis sample for the Washington state Measure of Student Progress (MSP) included 196 treatment students and 205 control students in grades 6-8 who received one year of exposure to the combined interventions.

The sub-sample of students who were assigned to receive both *Phonics Blitz* and *Read to Achieve* included 97 students; 49 students were randomly assigned to the treatment group and 48 were randomly assigned to the control group. The corresponding analytic sample included 32 treatment students and 31 control students on the GMRT and both Woodcock Reading Mastery scores, and 37 treatment and 39 control students on the MSP.

The sub-sample of students who were assigned to receive only *Read to Achieve* included 350 students; 173 were randomly assigned to the treatment group and 177 were assigned to the control group. The corresponding analytic sample included 144 treatment students and 151 control students on the GMRT, 143 treatment students and 151 control students on the Woodcock Reading Mastery scores, and 159 treatment and 166 control students on the MSP.

The Final Evaluation Report, which includes findings from the first year of implementation of the combined interventions, includes a *full* sample of students (students who received *Read to Achieve* and *Phonics Blitz* and students who received only *Read to Achieve*) large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .21 on the GMRT after one year of the interventions for grades 6-8,⁵
- .22 on the Woodcock Reading Mastery Word Identification subtest after one year of the interventions for grades 6-8,
- .22 on the Woodcock Reading Mastery Word Attack subtest after one year of the interventions for grades 6-8, and
- .22 on the MSP after one year of the interventions for grades 6-8.

Findings from the first year of implementation of the combined interventions for the sample of students assigned to receive both *Phonics Blitz* and *Read to Achieve* are based on a sample of students large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .67 on the GMRT after one year of the interventions for grades 6-8,⁶
- .60 on the Woodcock Reading Mastery Word Attack subtest after one year of the interventions for grades 6-8

⁵ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

⁶ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

- .62 on the Woodcock Reading Mastery Word Identification subtest after one year of the interventions for grades 6-8, and
- .51 on the MSP after one year of the interventions for grades 6-8.

Findings from the first year of implementation for students assigned to receive only *Read to Achieve* are based on a sample of students large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- .23 on the GMRT after one year of the intervention for grades 6-8,⁷
- .27 on the Woodcock Reading Mastery Word Identification subtest after one year of the intervention for grades 6-8,
- .26 on the Woodcock Reading Mastery Word Attack subtest after one year of the intervention for grades 6-8, and
- .25 on the MSP after one year of the intervention for grades 6-8.

Key Measures of Student Reading Outcomes (Source)

Gates-MacGinitie Reading Tests (GMRT) (External Test Publisher)

Woodcock Reading Mastery Test-Revised: Word Attack and Word Identification (External Test Publisher)

Measure of Student Progress (MSP) (State Test)

Implementation Study

Research Design and Methods: The fidelity of implementation of the intervention model was studied in the five treatment schools for the first year of implementation. The study assessed the fidelity of implementation of the professional development model and of the classroom instruction model. For each model, fidelity was defined as the extent to which the actual level of implementation measured during the study was consistent with the model as planned. For the fidelity of implementation of the instructional supports and professional development opportunities, scores were assigned attendance at the six types of professional development based on the attendance logs. Fidelity of implementation of the professional development model was rated as high, medium, or low, based on the level of participation of teachers in the training activities that were offered by the district. The definitions of the levels of fidelity were established by the evaluators and state project staff. Based on the percentage of the 90 hours of professional development attended, fidelity was rated as high (≥ 90 percent attendance), medium (70-89 percent attendance) or low (< 70 percent attendance).

In-class support for teachers was also scored as the percentage of in-class support teachers received divided by 14 (the minimum number of intended visits by a coach to each teacher). Fidelity of implementation of the in-class support was rated as adequate or not, based on the average hours of coaching received. Fidelity of implementation ratings were determined in consultation with program staff members and reflect only the quantity of in-class support received. Fidelity of implementation of the in-class support was rated as high, medium, or low,

⁷ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

with the definitions of the levels of fidelity established by the evaluators in consultation with the state project staff. High fidelity was defined as the teacher being visited by the coach at least 12 times, or ≥ 86 percent of intended coaching; medium fidelity was defined as the teacher being visited by the coach 8-11 times, or ≥ 57 -85 percent of intended coaching; and low fidelity was defined as the teacher being visited fewer than eight times, or < 57 percent of the intended coaching.

To measure fidelity of implementation of the classroom model, separate observation protocols were developed for *Phonics Blitz* and *Read to Achieve*. The *Phonics Blitz* observation protocol included 50 descriptors across seven program components: oral reading, phonemic awareness, phonics, word sort, detective work, words to read, and sentences to read. The descriptors were program-specific operations such as: teacher uses correct error procedures; students always use fingers when stretching sounds; teacher states objective. The rating scale for each descriptor was: (1) not very true of this lesson; (2) somewhat true of this lesson; and (3) very true of this lesson. The *Read to Achieve* protocol used three holistic rubrics to score fidelity of teacher activities and routines (ratings of 1-5), level of support (ratings of 1-3), and error correction (ratings of 1-3). These rubrics were applied to all lesson components: comprehension, vocabulary, comprehension with vocabulary, fluency, higher order thinking, and beyond the book. To calculate the fidelity ratings, the number of points given to the observation was divided by the number of possible points. The percentages were then translated into a rating of high fidelity (≥ 75 percent), medium fidelity (50-75 percent) or low fidelity (< 50 percent). (The cut points for these ratings were established in consultation with program authors and trainers.)

Our final question about implementation regarded the extent to which teachers completed all of the required lessons in *Read to Achieve* and *Phonics Blitz*. To examine this, we compared the number of lessons teachers reported completing to the number of lessons teachers were expected to complete. Data about lesson completion was collected twice. At the end of week 12, the point when teachers were supposed to be finished with *Phonics Blitz* in Group 1 classrooms, teachers reported through e-mail what lesson number(s) they had reached. At the end of the year, Striving Readers coaches reported this information for each teacher.

The study also calculated lesson completion by dividing the lessons teachers actually completed by the intended number of lessons. By the end of the year, *Phonics Blitz* teachers were expected to complete 50 lessons in 12 weeks, or just over four lessons per week. This was the standard for “high” lesson completion. For “medium” lesson completion, the cutoff was at least 43 lessons in 12 weeks while “low” completion was less than 43 lessons. By the end of the year, for *Read to Achieve* classes that also completed *Phonics Blitz*, “high” implementation meant completing 21 units from the content area curriculum, “medium” was 17-20 units and “low” was fewer than 17 units. For the classrooms that only implemented *Read to Achieve*, “high” implementation was defined as completing 30 units (21 units from the Content Area curriculum and the first nine units from the Narrative curriculum). According to the program developers, this matched a typical pace of four lessons per week. A “medium” level of completion was 25 units: 21 from the Content Area curriculum and at least four from the Narrative curriculum. Completion of fewer than 25 units represented a “low” level of completion.

Evaluation Findings

Fidelity of Implementation of the Intervention Model

In terms of the fidelity of implementation of the professional development, all six teachers had high levels of participation: five of the six teachers attended all 70 hours of professional development that was offered, and one teacher missed the first day of the summer institute, but attended all other trainings. In terms of in-class support, all teachers received at least 12 visits, and some received more. Support visits, which averaged 2 hours in length, amounted to 23 to 28 hours of coaching per teacher during the school year.

Based on multiple observations of the intervention classes, all teachers implemented both *Phonics Blitz* and *Read to Achieve* with high fidelity. Teachers had average fidelity scores between 77 and 92 percent. For *Phonics Blitz*, the overall level of fidelity of implementation was high for all teachers (with an average of 88 percent fidelity). Levels of fidelity of implementation for *Read to Achieve* were also high for five of the six teachers; the sixth teacher had an average rating of 67 percent, which translates to medium fidelity.

For lesson completion, fidelity of implementation was low. *Phonics Blitz* was designed to be covered in 12 weeks but took up to twice as long to teach. None of the teachers were able to complete the program by week 12; on average, they had only completed half of the intended lessons by week 12. After week 12, they continued teaching *Phonics Blitz* until all 50 lessons were taught, but this left them fewer remaining weeks to teach *Read to Achieve*. At the end of the year, teachers had completed between 29 and 57 percent of the *Read to Achieve* material that the program intended for them to cover in their classes. Lesson completion rates varied among classrooms where *Read to Achieve* was the only program taught all year. Two teachers had “high” lesson completion rates, covering all of the intended material by the last week of school; three teachers covered between 81 and 90 percent of the material (“medium” lesson completion); and one teacher covered 63 percent of the material (“low” lesson completion).

Impact of the Intervention on Student Reading Outcomes

For the full sample of students, there was a statistically significant impact on the reading achievement of struggling readers in grades 6-8 after one year of exposure to the combined interventions on the MSP. The effect size of the impact was .16. There was not a statistically significant impact on the reading achievement as measured by the GMRT or the Woodcock Reading Mastery Test. Effect sizes were .03, -.04, and .08 on the GMRT, and the Woodcock Word Identification and Word Attack subtests, respectively.

For the sample of students who received both *Phonics Blitz* and *Read to Achieve*, there were no statistically significant impacts on the reading achievement of struggling readers in grades 6-8 after one year of exposure to the combined interventions. The effect sizes were .16, .02, -.03, and .07 on the MSP, the GMRT, and the Woodcock Word Identification and Word Attack subtests, respectively.

For the sample of students who received only *Read to Achieve*, there were no statistically significant impacts on the reading achievement of struggling readers in grades 6-8 after one year of exposure to the combined interventions. The effect sizes were .11, .13, .14, and .33 on the

MSP, the GMRT, and the Woodcock Word Identification and Word Attack subtests, respectively.

Readers should exercise caution when drawing conclusions based on the lack of statistical significance of the findings. Because funding for the Striving Readers program was cancelled after one year of data collection, the study did not have an adequate sample size (i.e., number of students) to reliably detect some policy-relevant impacts of the effect of both *Phonics Blitz* and *Read to Achieve* or of only *Read to Achieve* on the reading outcomes of interest. The originally planned study design described in the grant application would have had a much larger sample size based on three years of data collection.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Intervention

Strengths:

- The analysis using the full sample of students as well as both subsamples of the impact of one year of *Phonics Blitz* and *Read to Achieve* on reading achievement (MSP, GMRT and Woodcock Word Identification and Word Attack subtests) meets What Works Clearinghouse (WWC) evidence standards. Analyses that meet WWC evidence standards make us the most confident that the effect we see is due solely to the intervention alone, and not to the many other factors that are at play in schools and in the lives of students, such as teachers, school and family.
- Eligibility for random assignment was determined systematically, using a predetermined cutoff score on a test of reading achievement (the Washington Assessment of Student Learning (WASL), the state reading assessment) and a measure of the number of words read correctly per minute.
- Random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.
- There was no evidence that other factors (e.g., other reading programs or district policies) were implemented in ways that would have undermined the evaluators' ability to attribute impacts to *Read to Achieve* or to *Phonics Blitz*.
- The evaluation employed three reading tests as outcome measures. The Gates-MacGinitie Reading Test (GMRT), a measure of reading comprehension; the Woodcock Reading Mastery Test-Revised, Word Attack and Word Identification subtests, measures of phonics; and the Measure of Student Progress (MSP), the Washington state test of general literacy achievement. There was no reason to believe that students assigned to the treatment group had more experience taking the test than did the control group students, or that the tests measured skills specific to the intervention, both of which could have undermined confidence in the impact estimates.

- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement was included in the models to increase the precision of the impact estimates.
- While some students were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggested that the integrity of the original randomized design was preserved, and that treatment and control groups continued to be statistically equivalent on all measured and unmeasured characteristics at follow-up.
- For the combined sample (students who received only *Read to Achieve* and students who received both *Phonics Blitz* and *Read to Achieve*):
 - 14.1 percent of grade 6-8 students who received one year of intervention were unable to participate in follow-up data collection on the GMRT. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 0.00 percent. This amount of attrition is within the acceptable range established by WWC standards.
 - 14.4 percent of grade 6-8 students who received one year of intervention were unable to participate in follow-up data collection on the Woodcock Reading Mastery Test (both subtests). The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 0.5 percent. This amount of attrition is within the acceptable range established by WWC standards.
 - 3.8 percent of grade 6-8 students who received one year of intervention were unable to participate in follow-up data collection on the MSP. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.1 percent. This amount of attrition is within the acceptable range established by WWC standards.
- For the sample of students who received both *Phonics Blitz* and *Read to Achieve*:
 - 21.2 percent of grade 6-8 students who received one year of intervention were unable to participate in follow-up data collection on the GMRT and on the Woodcock Reading Mastery Test. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 0.00 percent. This amount of attrition is within the acceptable range established by WWC standards.⁸
 - 5.0 percent of grade 6-8 students who received one year of intervention were unable to participate in follow-up data collection on the MSP. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 0.2 percent. This amount of attrition is within the acceptable range established by WWC standards.

⁸ For more information, please see Appendix A-Assessing Attrition Bias, of the WWC Procedures and Standards Handbook, available at: <http://ies.ed.gov/ncee/wwc/help/idocviewer/Doc.aspx?docId=19&tocId=7>.

- For the sample of students who received only *Read to Achieve*:
 - 12.5 percent of grade 6-8 students who received one year of intervention were unable to participate in follow-up data collection on the GMRT. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.6 percent. This amount of attrition is within the acceptable range established by WWC standards.
 - 12.8 percent of grade 6-8 students who received one year of intervention were unable to participate in follow-up data collection on the Woodcock Reading Mastery Test (both subtests). The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 2.2 percent. This amount of attrition is within the acceptable range established by WWC standards.
 - 3.6 percent of grade 6-8 students who received one year of intervention were unable to participate in follow-up data collection on the MSP. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 1.3 percent. This amount of attrition is within the acceptable range established by WWC standards.

Weaknesses:

- None.

Summary of 2009 Striving Readers Projects: Profile of Wisconsin Department of Public Instruction Striving Readers Project and Evaluation¹

Grantee: Wisconsin Department of Public Instruction

Project Director: Jacqueline Karbon

Local Evaluator: Learning Point Associates

Principal Investigator: Andrew Swanlund

Project Website: <http://www.wsra.org/archives/dpi0705.php>

Setting of the Study

The Wisconsin Striving Readers Program was implemented in five Title I schools in the Milwaukee Public Schools (MPS). In 2008-2009, these schools served 3,944 students. 70 to 96 percent of these students were eligible to receive free/reduced price lunch; an average of 40 percent did not meet the state standards in reading achievement. The Striving Readers project was funded for two years, which included one year of implementation of the intervention during the 2010-11 school year.

Intervention Model

Classroom Model as Planned²

The Wisconsin Striving Readers Program's READ 180 intervention was developed by Scholastic. The curriculum is designed to address individual student needs through adaptive instructional software, high interest literature, and direct instruction in reading and writing skills.

The READ 180 program follows a structured instructional model. Each class begins with twenty minutes of whole-group instruction. Students are then broken up into three small groups and spend twenty minutes in each of the following rotations: small-group instruction, instructional software, and modeled and independent reading. The class period ends with time for students to reflect on what they learned that day during a ten-minute, whole-group session.

The primary text used for instruction is the student rBook, an interactive text divided into nine instructional workshops. Each workshop contains six sections: (1) Preview/Teach Vocabulary; (2) Reading; (3) Vocabulary/Word Study; (4) Writing and Grammar; (5) Functional Literacy; and (6) Workshop Wrap-Up. Throughout each workshop there are "checkpoints" that allow teachers to assess student knowledge and provide opportunities for more in-depth skill instruction and practice. Teachers will make frequent use of assessment data created by the

¹ Swanlund, A., Dahlke, K., Tucker, N., Kleidon, B., Kregor, J., Davidson-Gibbs, D., Hallberg, K. *Striving Readers: Impact Study and Project Evaluation Report*. (2012). Naperville, IL: American Institute for Research.

² For more information on *READ 180*, please see the READ 180 Intervention Profile by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

READ 180 software system to identify the individual strengths and weaknesses of students, thus allowing them to tailor instruction to the individual needs of their students.

Professional Development Model as Planned

The Wisconsin Striving Readers intervention was delivered by eight intervention teachers – one teacher at each of seven schools and an eighth teacher who split her time between two schools. Four types of professional development opportunities were intended to serve as resources for implementing the program. Scholastic provided initial READ 180 implementation training, including training on strategies to differentiate instruction for students with disabilities and English Language Learners (ELLs); use of data to assess student achievement and inform instruction; traits-based writing strategies; classroom management and design; and software management instruction. The professional development component began with an initial three-day group session. The first two days of training introduced the program components and demonstrated how READ 180 was designed to address individual needs through small group instruction, computer software, and high interest literature. The third day of the training provided review on classroom management and goal setting skills.

The model called for intensive ongoing professional development sessions that would provide specific curriculum implementation training as well as best practices in literacy and data analysis. The ongoing training was delivered via the Scholastic online RED courses. These courses were designed to increase teacher knowledge about adolescent literacy. RED course lessons guide the teacher through practice and real-world application. In Year 1, the online course topics included “Best Practices for Reading Intervention” (for novice READ 180 teachers) and “Strategies for Teaching Striving Readers” (for all READ 180 teachers). Participating in the RED courses was required. These courses spanned 14 sessions, with each session lasting half an hour. The READ 180 teachers were also expected to participate in eight monthly roundtable sessions to discuss what they learned in the RED courses and to provide teachers with the opportunity to ask questions and share ideas with one another. These sessions were planned to last six hours each. As part of their professional development training, all READ 180 teachers were trained to use the Scholastic Achievement Manager (SAM) data system. Delivering these various group trainings was a collaborative effort on the part of Scholastic, the Wisconsin Department of Public Instruction, and Milwaukee Public Schools.

To try to ensure that reading intervention teachers were highly qualified, Carroll University offered graduate-level coursework, which was required for teachers without a reading license. Teachers could earn credits toward their reading license, and teachers who already had a reading license were to be granted credits toward their reading specialist license. Tuition reimbursement of up to 12 credits was offered to teachers who earned a ‘B’ or better.

Another form of ongoing professional development provided to the intervention teachers was mentoring from more experienced teachers and coaches. A coach from Scholastic was expected to conduct observations of each READ 180 intervention teacher once a month for one hour and provide feedback. The project coordinator from MPS also was expected to conduct observations and provide feedback as needed. In addition, District Identified for Improvement (DIFI) supervisors or literacy coaches were assigned to spend time in the classrooms observing, providing feedback, and modeling appropriate instructional strategies as needed.

Context for Implementation

Because the main focus of this study was to estimate the effect of receiving two years of a supplemental READ 180 intervention, students randomized into the treatment group were placed in READ 180 classrooms, where they were expected to participate for two consecutive years. In the first year of implementation, the Wisconsin Striving Readers program identified approximately 800 eligible 6th-9th grade struggling readers to participate in the study, with half of those eligible students randomly assigned to the treatment group and half to the control group. For students in the treatment group, READ 180 was implemented as a supplement to regular English Language Arts (ELA) classes, taking the place of an elective for a 90-minute block each day. Students were assigned to the treatment group to receive the intervention for two years, but because Congress eliminated the Striving Readers program mid-way through the grant, the study only followed students through one year of the intervention.

Students were eligible to participate in Wisconsin Striving Readers if they received a score of “Minimal” or “Basic” on the Wisconsin Knowledge and Concepts Examination (WKCE). Students who did not have a WKCE score were also eligible based on having a score of “Minimal” or “Basic” on the Education Assessment Predictive Benchmark Assessment (ThinkLink). If a student did not have a recorded score for either of these assessments, eligibility for the intervention could also be established based on teacher assessments and observations that indicated that the student (i.e., a rating of 1 on the 1-4 scale of the Student Promotion System) was performing at least two grade levels below expectations.

Students with disabilities who were classified as Language Level A in the fall of each year and students who had not completed the Language course for an entire year were not included in the study. In addition, students with disabilities participating in Unique (students in this program take the state’s alternative assessment) were excluded from participation. English Language Learners below LAU Level 3 were excluded from the sample, as were ELLs for whom the intervention would interfere with their regular ESL services. Lastly, students who were deemed to be proficient or advanced by a principal or teacher also could be excluded.

Evaluation Design

Research Questions

1. What is the impact of one year of READ 180 participation on struggling 6th-9th grade students’ academic performance in reading?
2. What is the fidelity of implementation of READ 180 in the study classrooms?

Impact Study

Research Design and Methods: The impact of one year of the intervention on students’ reading achievement and engagement was estimated using an experimental design and within-school random assignment of students. Models were run for each outcome and included schools as fixed effects, pretest scores on each outcome measure, and additional student-level covariates, including gender, ethnicity, economic status, special education status, and ELL status.

Control Condition: Students randomly assigned to the control group attended a study hall or elective class. Students in the control group attended regular English Language Arts (ELA) classes along with their classmates in the treatment group. At the time of the study, no systematic support services existed for struggling readers as part of their regular English Language Arts (ELA) classes. Any such support that did exist, however, would have been available to both treatment and control students. Any supplemental services being provided to or sought out by students involved in the study were monitored by the evaluators. The student engagement survey, given to both treatment and control students, also asked students to indicate any supplemental support services they received during the study year.

Sample Size: The original study focused on measuring the impact of a two-year reading intervention on students in grades 6-10. Because the Striving Readers grant was discontinued, this study will focus on the impact of a one-year reading intervention on students in grades 6-9. Approximately 900 struggling readers in grades 6-9³ from five schools in the MPS were eligible to participate in the study in the first year of implementation of the intervention. Of those, 462 students were randomly assigned to the treatment group, and 419 were assigned to the control group (the remaining students were placed on a waitlist). For the Northwest Evaluation Association Measures of Academic Progress (MAP), the impact analysis included 335 treatment and 284 control students in grades 6-9. The sample of students for the impact analysis is large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to .19 on the MAP after one year of the interventions for grades 6-9.⁴

Key Measures of Student Reading Outcomes (Source)

Northwest Evaluation Association Measures of Academic Progress (MAP) (External Test Publisher)

Implementation Study

Research Design and Methods: The fidelity of implementation of the intervention model was studied in the five treatment schools for the first year of implementation. The study assessed the fidelity of implementation of the *professional development model* and of the *classroom instruction model*. For each model, fidelity was defined as the extent to which the actual level of implementation measured during the study was consistent with the model as planned. For the fidelity of implementation of the instructional supports and professional development opportunities, scores were assigned attendance at the six types of professional development, based on the attendance logs. Fidelity of implementation of the professional development model was rated as adequate or not, based on teachers' level of participation in the training activities offered by the district. The total number of points that could be awarded was 51, with a low level of implementation corresponding to 28 points or less, a medium range corresponding to 29-40 points, and a high level of implementation corresponding to 41 points or higher.

³ If the study had continued for two years, the intervention would have been extended into 10th grade, for the students who started out in 9th grade in the first year.

⁴ Abt Associates staff calculated the MDE by multiplying the standard error of the impact estimate by 2.8. This calculation produces the MDE for a two-tailed test with 80% power, and with an alpha level of .05, and accounts for clustering and for the inclusion of the covariates in the model.

To measure fidelity of implementation of the classroom model, the study used multiple data sources (classroom observations, interviews, and extant program data) to examine the extent to which the program provided instruction to students in accordance with the intervention logic model.⁵ The evaluator developed a rubric to rate components of READ 180, which included class size, student attendance, and classroom model components (whole-group instruction and skills lesson for 20 minutes, rotations between small-group instruction (computer, small group, and modeled and independent reading rotations) each lasting 20 minutes, and a 10 minute wrap-up session during which the teacher facilitates discussion about the day's lesson). Each classroom was rated as low, medium, or high implementation, where low is 35-58 points, medium is 59-82 points, and high is 83-105 points.

Evaluation Findings

Fidelity of Implementation of the Intervention Model

In terms of fidelity of implementation of the *professional development model*, in Year 1, five of eight classrooms received a score in the medium range and three classrooms received a score in the high range (41-51). The average score across all classrooms was 39 (a medium level of fidelity). Some elements of the professional development model received more ratings in the low range than did other items. For example, all classrooms received high ratings for the orientation training and all classrooms received low ratings on the mentoring support.

In terms of fidelity of implementation of the *classroom model*, seven of eight classrooms received a rating in the medium range and one classroom received a score in the low range. With some exceptions, teachers implemented the READ 180 with high fidelity, but low student attendance reduced the overall classroom fidelity scores.

Impact of the Intervention on Student Reading Outcomes

There was a statistically significant impact on the reading achievement of struggling readers in grades 6-9 after one year of exposure to READ 180. On the MAP, the effect size was .14.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Intervention

Strengths:

- The analysis of the impact of one year of READ 180 on reading achievement (MAP) meets What Works Clearinghouse (WWC) evidence standards. Analyses that meet WWC evidence standards make us the most confident that the effect we see is due solely to the intervention alone, and not to the many other factors that are at play in schools and in the lives of students, such as teachers, school and family.

⁵ The definition of fidelity of implementation of the classroom model for the MCLA whole school intervention can be found on the website for the U.S. Department of Education, Office of Elementary and Secondary Education in the cross-site summary tables for the 2006-2011 cohort (*Summary of 2006 Striving Readers Projects. Implementation and Evaluation of Targeted Interventions for Struggling readers and Whole School Interventions for All Readers: Years 1 – 5*), Table 15.

- Eligibility for random assignment was determined systematically, using a predetermined cutoff score on a test of reading achievement (Wisconsin Knowledge and Concepts Examination (WKCE) or ThinkLink (the district benchmark assessment)), or teacher evaluations of reading proficiency conducted prior to random assignment.
- Random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.
- There was no evidence that other factors (e.g., other reading programs or district policies) were implemented in ways that would have undermined the evaluators' ability to attribute impacts to READ 180.
- The evaluation employed one reading test as an outcome measure, the Northwest Evaluation Association Measures of Academic Progress (MAP), which is a measure of reading comprehension. There was no reason to believe that students assigned to the treatment group had more experience taking the test than did the control group students, or that the tests measured skills specific to the intervention, both of which could have undermined confidence in the impact estimates.
- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement was included in the models to increase the precision of the impact estimates.
- While some students were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggested that the integrity of the original randomized design was preserved, and that treatment and control groups continued to be statistically equivalent on all measured and unmeasured characteristics at follow-up.
 - 29.7 percent of grade 6-9 students who received one year of READ 180 were unable to participate in follow-up data collection on the MAP. The level of attrition did not differ substantially across the treatment and control groups; the differential attrition rate was 4.7 percent. This amount of attrition is within the acceptable range established by WWC standards.⁶

Weaknesses:

- None.

⁶ For more information, please see Appendix A-Assessing Attrition Bias, of the WWC Procedures and Standards Handbook, available at: <http://ies.ed.gov/ncee/wwc/help/idocviewer/Doc.aspx?docId=19&tocId=7>.