

Summary of 2006 Striving Readers Projects: Profile of Newark Public Schools' Striving Readers Project and Evaluation

Grantee: Newark Public Schools
Project Director: Gayle Griffin, Ph.D.
Local Evaluator: Westat
Principal Investigator: Jennifer Hamilton

Setting

Nineteen middle schools participate in the Newark Striving Readers Project. These schools were selected because they 1) were eligible for Title I funding, 2) served a minimum of two grades across grades six through eight, 3) were not already using READ 180, 4) were categorized as “in need of improvement” under No Child Left Behind, and 5) served a minimum of 25 eligible students reading at least two grades below grade level, based on the 2006 New Jersey state assessment. In these schools, 58 percent of students are African American, 41 percent are Hispanic, 88 percent are identified as low-income, and 7 percent are identified as being English Language Learners (ELL).

Intervention Models

Targeted Intervention

Classroom Model as Planned: The READ 180 program, developed by Scholastic Inc, aims to address the individual needs of struggling adolescent readers who are reading below grade level through adaptive and instructional software, teacher-directed instructional rotations, and the use of tailored textbooks and independent or modeled reading of literature intended to be of high interest to adolescents. The Newark Striving Readers project modified READ 180 to include some supplemental instruction aligned with the state assessment. The program focuses on elements of phonics, fluency, vocabulary, comprehension, spelling, writing and grammar, and aims to promote self-directed learning. Daily assessments are provided by the READ 180 Topic Software and the Scholastic Achievement Manager (SAM) software provides feedback to teachers on student assessments. In addition, diagnostic testing using the Scholastic Reading Inventory (SRI) is conducted three times a year.¹

Professional Development Model as Planned: In the first year of the study, teachers are offered an initial two-day training on the model with a follow-up session on use of data and an individual conference, all provided by the model developer. These training modalities comprise about 18 hours of professional development. In the second year of the study, teachers (those who were new to READ 180 and returning teachers) are offered a one-day initial training and a follow-up session. In the third year, professional development was offered only to new teachers, who are offered the one day initial training only. In all three years, teachers are offered in-class

¹ For more information on READ 180, please see the READ 180 Enterprise Edition Intervention Profile by Abt Associates, available at <http://www2.ed.gov/programs/strivingreaders/performance.html>.

technical assistance from the district Resource Teacher Coordinator as needed. In the first year of the study, teachers also receive classroom visits from a READ 180 coach on an as-needed basis.

Context for Implementation: The READ 180 program is being implemented in 10 middle schools, as a replacement of the district's regular core language arts curriculum. Both models are being implemented as supplements to the regular English language arts (ELA) curriculum in the schools. Students in grades 6-8 are eligible for READ 180 if they score at least 1 standard deviation below the grade-standardized mean on the reading subtest of the New Jersey state assessment (ASK: Assessment of Skills and Knowledge). Eligible students can receive the interventions for up to three years (grades 6-8). All special education students who are struggling readers are eligible for the interventions. In the first year of the implementation, approximately 700 students were assigned to READ 180 classes in the 10 treatment schools. In the second year of implementation, a new cohort of 6th graders was added. The number of students served in Year 2 of implementation was approximately 600, including new 6th graders and students in grades 7 and 8 who were eligible to continue in READ 180 for a second year. The targeted intervention will be implemented for a total of four years.

Whole School Intervention

Classroom Model as Planned: The whole school intervention combines two professional development programs on literacy instruction from two providers, New Jersey City University (NJCU) and the National Urban Alliance (NUA). NJCU provides the professional development for language arts teachers and literacy coaches, helping teachers guide students in using a variety of strategies for helping students comprehend text, such as graphic organizers, text annotation, note-taking, post-reading reflection, and anticipation guides to model brain-storming. NUA provides the professional development for other content teachers, to help them provide instruction in similar strategies for content lessons, such as graphic organizers, anticipation guides, and word taxonomies.

Professional Development Model as Planned: Prior to their first year of implementing either of the whole school models, teachers are offered initial Summer Institutes lasting either three or four half-days, for training by NUA or NCJU, respectively. Language arts teachers also have three follow-up training sessions with NCJU trainers during the year and NCJU coaches visit each of the 19 schools five times a year, for observation and discussion with teachers. Other content area teachers have two follow-up training sessions with NUA during the year, and NUA coaches visit each school 15 times to observe and work with teachers. District resource teacher coordinators are also available to visit schools to provide technical assistance on the whole school model on an as-needed basis. In subsequent years, the professional development model delivered by NCJU to the language arts teachers increases in intensity with the addition of three follow-up training sessions. The professional development for the other content area teachers remains the same for all three years..

Context for Implementation: All teachers of language arts, mathematics, science, social studies, and special education in the 19 participating middle schools are trained on the whole school interventions. All students in the participating schools are taught by teachers who are part of the model. In a single school year, this involves approximately 3,600 students in grades 6-8

across the 19 schools. The whole school intervention will be implemented for a total of four years.

Evaluation Design

Evaluation of the Targeted Interventions

Research Questions:

1. Does READ 180 significantly improve the reading skills of targeted students?
2. Do different types of students benefit from the intervention in different ways?
3. Does READ 180 significantly improve the school attendance of targeted students?

Research Design and Methods: School-wide random assignment was used to assign the 19 participating middle schools to either implement READ 180 or to continue to provide only the regular language arts curriculum. Schools were blocked on four school-level variables prior to randomization: 1) school size, 2) proportion of ELL students, 3) proportion of students with special needs, and 4) AYP status.

The impact of READ 180 on student outcomes will be assessed using multilevel models to account for the nesting of students within schools. Analyses will also be run to disaggregate effects by grade level, gender, ELL, and special education.

Control Condition: Students in schools randomized to the control condition continue to receive instruction in the regular language arts curriculum.

Sample Size: Nineteen schools were randomly assigned—ten to implement READ 180 and nine to the control condition. In the first year of the study, 1,371 students participated in the evaluation, across both treatment conditions. A new cohort of approximately 400 6th grade students was added in the second year of the study, 200 in treatment schools and 200 in control schools. The impact analysis was conducted on a sample of 934 students in grades 6-8 who had been in READ 180 for one year in either the first or the second years of implementation. The sample size for the control group was 838 students in nine middle schools. A third cohort of students was added in the third year of the study, such that the analysis of the impact of one year of READ 180 was conducted on a sample of 1,149 treatment students and 1,022 control students. The analysis of the impact of two years of READ 180 was conducted on a sample of 624 treatment students and 545 control students and the analysis of the impact of three years of READ 180 was conducted on a sample of 207 treatment students and 182 control students.

The Year 3 evaluation report, which includes findings from the first three years of implementation, includes a sample of students large enough to detect an impact (in standard deviation units) of the intervention on student outcomes equivalent to:

- .12, .14, and .12 on the standardized test (SAT 10) of vocabulary, comprehension, and language arts respectively, for students in grades 6-8 who received one year of READ 180,²
- .16, .20, and .15 on the standardized test (SAT 10) of vocabulary, comprehension, and language arts respectively, for students in grades 6-8 who received two years of READ 180,
- .20, .18, and .17 on the standardized test (SAT 10) of vocabulary, comprehension, and language arts respectively, for students in grades 6-8 who received three years of READ 180, and
- .01 on the attendance measure, for students in grades 6-8 who received one to three years of READ 180.

Because Newark will continue to offer the intervention to new groups of students for four school years, the Year 4 report will have larger sample sizes and be able to detect smaller impacts, particularly in the analyses of the effects of multiple years of READ 180. However, because the Newark evaluation conducted random assignment at the school level, the additional gains in precision will be minimal.

Key Measures of Student Reading Outcomes (Source):

Stanford Achievement Test-10 (vocabulary, reading comprehension, and language arts subtests)
(External Test Publisher)

Evaluation of the Whole School Intervention

Research Questions:

1. Does participation in an ongoing literacy professional development program change the instructional practices of middle school teachers?
2. Does participation in an ongoing literacy professional development program affect the instructional practices of some groups of teachers more than others?
3. Do these changes in teacher instructional practices result in improved reading skills of middle school students?

Research Design and Methods: An interrupted time series analysis will be used to compare pre-program student reading proficiency scores with post-program student reading proficiency scores on the New Jersey state proficiency test.

Future evaluation reports will include findings on the impact of the whole school intervention on student achievement. The interrupted time series evaluation design is made more rigorous with the inclusion of more than two years of post-implementation data.

Comparison Group: All schools in the study participate in the whole school intervention. Therefore, there is no comparison group.

² 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.

Sample Size: Approximately 480 teachers from the 19 participating schools were eligible to participate in the whole school intervention. All students in the 19 schools in each year in the time series will be included in the final evaluation of the whole school intervention. This consists of approximately 3,600 students. Students in these schools are compared to students in all other middle schools in the districts that serve students in grades 6-8 (24 schools for grade 6, 21 schools for grade 7, and 20 schools for grade 8).

The Year 3 evaluation report, which includes findings from the first three years of implementation, includes a sample of students large enough to detect an impact (in standard deviation units) of the intervention on reading achievement equivalent to:

- Grade 6, one year exposure--.65, two years exposure--.67
- Grade 7, one year exposure--.56, two years exposure--.56
- Grade 8, one year exposure--.44, two years exposure--.44.

Key Measures of Student Reading Outcomes (Source):

New Jersey State Language Arts assessment (State Test)

Year 3 Evaluation Findings

Evaluation of the Targeted Intervention

Fidelity of Implementation of the Targeted Intervention Model: In terms of fidelity of implementation of the *professional development model*, in Year 1 of implementation, the majority of teachers participated in the READ 180 professional development activities at either an adequate level (22%) or fully (56%). The level of participation in professional development declined in Year 2 of implementation, with 24% of teachers participating at an adequate level and 8% participating fully. No additional training was offered to teachers in Year 3.

In terms of fidelity of implementation of the *classroom model*, ratings of the classrooms were based on a combination of administrative data and data from the READ 180 computerized systems. The student assessments were implemented in line with the READ 180 guidelines in nearly all of the classrooms in Years 1, 2, and 3 of the program. In Year 1 (2006-07), students spent an adequate amount of time using the READ 180 instructional software (compared to the program guidelines) in nearly all of the classrooms (91%). In Year 2 (2007-08), this proportion dropped to fewer than 10% of the classrooms, and in Year 3 (2008-09), this proportion was 50% of the classrooms.

Impact of the Targeted Interventions on Student Reading Outcomes: For treatment students who had one year of READ 180, there were no significant effects on any of the three subtests of the Stanford Achievement Test. The effect sizes for the three subtests, vocabulary, comprehension, and language arts, were .09, .10, and .07, respectively.

For treatment students who had two years of READ 180 there were significant effects on two of the three subtests of the Stanford Achievement Test. The effect sizes for vocabulary and comprehension were .09 and .17, respectively. No significant effects were found on the language arts subtest; the effect size was .10.

For treatment students who had three years of READ 180 there were significant effects on one of the three subtests of the Stanford Achievement Test. The effect size for this subtest, language arts, was .17. No significant effects were found on the vocabulary or comprehension subtests. The effect sizes were .07 and .14, respectively.

For treatment students who had one, two, or three years of READ 180 there were no significant effects on attendance with an effect size of .00 in all three years.

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

Strengths

- Eligibility for random assignment was determined systematically, using a predetermined cutoff score on a test of reading achievement (reading subtest of the New Jersey Assessment of Skills and Knowledge (NJ-ASK)).
- School-level random assignment was faithfully executed, with no evidence of students receiving the intervention after being randomized to the control condition.
- There is no evidence that there are other factors (e.g., other reading programs or district policies) that were implemented in ways that would undermine the evaluators' ability to attribute impacts to READ 180.
- When estimating impacts, appropriate analytic steps were taken to account for the clustering of students within schools. A pre-study measure of reading achievement is included in the models to increase the precision of the impact estimates.
- Assessment of Student Outcomes:
 - The reading test used as an outcome measure (SAT-10) assesses language arts, comprehension, and vocabulary, and was developed by an external test publisher. There is no reason to believe that students in schools assigned to the treatment group have more experience taking these tests than do the students in schools assigned to the control group, or that the test measures skills specific to the intervention, both of which could undermine confidence in the impact estimates.
 - The measure of student attendance is extracted from district student records. Since it is measured the same way for all students, there is no reason to believe that the measurement undermines confidence in the impact estimates.
- All schools were able to participate in follow-up data collection in Year 3. While some students within schools were unable to participate in follow-up data collection, the level of attrition did not differ substantially across the treatment and control groups. This suggests that the integrity of the original randomized design was preserved, and that the treatment and control groups continue to be equivalent on all measured and unmeasured characteristics at follow-up.
 - Some students (7.6% of grade 6-8 students) were unable to participate in follow-up data collection on the SAT10 after one year of treatment; the levels of attrition did not

- differ substantially across the treatment and control groups (differential attrition rate was 1.8%). This amount of attrition is within the acceptable range established by WWC standards.³
- Some students (14.5% of grade 7-8 students) were unable to participate in follow-up data collection on the SAT10 after two years of treatment; the levels of attrition did not differ substantially across the treatment and control groups (differential attrition rate was 4.8%). This amount of attrition is within the acceptable range established by WWC standards.
 - Some students (25.3% of grade 8 students) were unable to participate in follow-up data collection on the SAT10 after three years of treatment; the levels of attrition did not differ substantially across the treatment and control groups (differential attrition rate was 4.1%). This amount of attrition is within the acceptable range established by WWC standards.
- Despite random assignment and low attrition, differences in treatment and control schools in eligibility for free and reduced priced lunch were noted. The effects of these differences are mitigated by the inclusion of this measure in the statistical models estimating the impact of the program.

Weaknesses

- None.

Evaluation of the Whole School Interventions

Fidelity of Implementation of the Whole School Intervention Model: In terms of fidelity of implementation of the *professional development model*, in Year 1 of implementation, across the two models (NUA and NJCU), in 42% of the schools, teachers participated at a moderate-to-high level in the whole school professional development, but there were no schools where the teachers participated at a high level. In Year 2, there were again no schools in which teachers participated at a high level and the percentage of schools in which teachers achieved moderate-to-high participation dropped to 16%. In Year 3, 5% of schools achieved high levels of teacher participation in professional development and 47% of schools achieved moderate-to-high levels of participation.

With respect to the implementation of *in-class support or coaching*, in the first year, in 68% of schools teachers received a high level of coaching support and in another 32% of schools teachers received a moderate-to-high level of coaching support. These percentages increase over the next two years of the study. In the second year, in 89% of schools teachers received a high level of coaching support and in all of the remaining schools teachers received a moderate-to-high level of coaching support. In the third year, teachers received a high level of coaching support in 100% of the schools.

³ 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/iddocviewer/Doc.aspx?docId=19&tocId=7>.

Impact of the Whole School Intervention on Student Reading Outcomes: There were no statistically significant impacts of the whole school intervention on student reading outcomes after one or two years of the intervention. The non-significant impacts after one year were .23, .20, and .16 for grades 6, 7, and 8 respectively. The non-significant impacts after two years were .24, .20, and .16 for grades 6, 7, and 8 respectively.

Summary of the Strengths and Weaknesses of the Impact Evaluation of the Whole School Intervention:

Strengths

- *Comparison Group.* A comparison group of schools was constructed by selecting all other schools in the same districts as the Striving Readers schools that serve students in at least one of the middle grades (6, 7, or 8). This method of selecting a comparison group attempts to establish a group of schools that is similar to the Striving Reader schools in terms of student achievement and demographic characteristics. Nevertheless, we cannot assume that students in treatment and comparison schools are alike in all characteristics, observed and unobserved, that could affect their reading achievement.
- *Baseline Equivalence:* The achievement levels, as measured by the Language Arts Literacy (LAL) section of the NJASK, of students in Striving Readers schools and the comparison schools were statistically equivalent (at each grade level, 6, 7 and 8) prior to the implementation of the Striving Readers Program. This is consistent with WWC evidence standards, with reservations.
- *Outcome Measure:* The reading test used as an outcome measure (NJASK-LAL) assesses language arts literacy, including reading and writing skills and was developed by the state. There is no reason to believe that students in schools assigned to the treatment group have more experience taking these tests than do the students in schools assigned to the control group, or that the test measures skills specific to the intervention, both of which could undermine confidence in the impact estimates.
- *Attrition:* All 19 Striving Readers schools along with the complete sample comparison schools in the same school districts were included in the analysis.
- *Analysis:* When estimating impacts, appropriate steps were taken to account for the clustering of students within schools, and pre-existing differences in reading achievement.

Weaknesses

- *Presence of Confounding Factors:* There is no evidence that there are other factors (e.g., other reading programs or district policies) that were implemented in ways that would undermine the evaluators' ability to attribute impacts to the whole school intervention. However, they may be pre-existing differences between the two groups that might be responsible for any observed impacts. Therefore, we are unable to attribute observed impacts to the whole school intervention.

- *Limitation.* The impacts of the whole school intervention for struggling readers cannot be estimated independently from the targeted intervention because some students in the ten treatment schools also receive READ 180 instruction.