RESEARCH REPORT

Effectiveness of Scholastic's *READ 180* as a Remedial Reading Program for Ninth Graders: A Report of an Implementation in Anaheim, CA

Denis Newman
Empirical Education Inc.

Andrew Jaciw
Stanford University

May 1, 2006

Empirical Education Inc.
www.empiricaleducation.com
425 Sherman Avenue, Suite 210
Palo Alto, CA 94306
(650) 328-1734

Acknowledgements

We are grateful to the Anaheim Union High School District for providing the data for this analysis through its agreement with Scholastic, Inc., which sponsored the research.

About Empirical Education Inc.

Empirical Education Inc. was founded to help K–12 school districts, publishers, and the educational R&D community assess new or proposed instructional programs through scientifically based pilot implementations. The company draws on the expertise of world-class researchers and methodologists assuring that the research is objective and takes advantage of current best practice in rigorous experimental design and statistical analysis. The company's findings let educators quantify the value of programs and help them partner with providers to implement those most effective for their students.

Reference this report: Newman, D., & Jaciw, A. (2006, May). Effectiveness of Scholastic's READ 180 as a remedial reading program for ninth graders: Report of an implementation in Anaheim, CA. Palo Alto, CA: Empirical Education Inc. Retrieved 2020 from https://www.empiricaleducation.com/past_research/

Executive Summary

Scholastic, Inc. contracted with Empirical Education Inc. to work with the Anaheim Union High School District to evaluate the impact of its *READ 180* program in improving reading performance for 9th graders for the 2004-2005 school year. Our research question was whether students in classes using *READ 180* improved more than expected on measures of reading. Without a rigorously designed control condition, our study could not address causal relationships; nevertheless, our approach provides a useful metric for assisting the district in evaluating their initial implementation. Within the limits of the current study, the research supports the district's decision to implement *READ 180*.

Intervention. *READ 180* is complex program with a specific implementation plan. The expectation is that reading classes are organized as 90-minute periods that are structured into a 20-minute whole group segment; 60 minutes of rotation through small group instruction, independent reading, and computer use; and a 10-minute wrap-up. The materials include interactive worktexts, paperbacks for independent reading, individualized instructional software, and audiobooks for modeled reading.

Setting. Anaheim UHSD serves more than 32,000 middle- and high-school students: 55% Hispanic, 25% White, 11% Asian, and 3.4% Black. The district has organized reading classes for students who read below grade level and are not English language learners. Overall the district scores slightly higher than the California average in English Language Arts proficiency.

Research design. Our study is a pre- to post-intervention comparison of performance for students using *READ 180* during one school year. Although there was no control or comparison group, there was an expectation that current students would attain one year of growth on the Gates-MacGinnitie Reading Test (GMRT). We were to determine whether there was a significantly greater or lesser amount of gain in GMRT performance than what was expected.

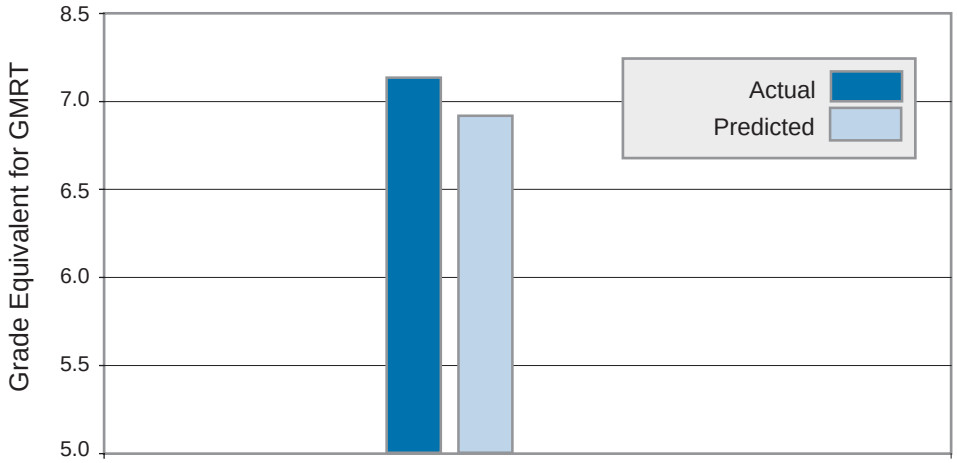
Participants. Students scoring below the 40th percentile on the California Achievement Test (CAT/6) or more than two levels below grade on the GMRT were eligible for a reading class that used *READ 180* or other interventions. Assignment was determined by the availability of *READ 180* licenses. Thus the actual range for scores of students receiving *READ 180* differed somewhat from school to school, depending on availability. Our study involved selected schools and teachers in which *READ 180* was adequately implemented. After attrition, the sample consisted of 11 teachers in six schools and their classes, which included 537 students.

Implementation. The school year 2004-2005 was the district's second year of implementation for *READ 180*. Five of the 11 teachers had already used the program for a full year, two had begun half-way through the previous year, and for four, this was the first year of implementation. All the classes considered were at the 9th grade level and all were using the high-school version of the product.

Although classes met each day, none of the schools utilized the expected 90-minute class periods; and only three of the teachers reported using the recommended full sequence of whole-class instruction, rotation, and wrap-up. Others used shorter periods, generally dropping the whole-class instruction at the beginning of class. All the classrooms were well equipped with the books and other equipment that are an integral part of the program. Because of the very small numbers of teachers, we did not attempt to draw any conclusions as to possible impact of these implementation differences on student achievement.

Statistical analysis. The district provided us with student test score data. The pretest and outcome measures we used were student scores on the GMRT, which was given in September and March, yielding a six-month difference between pre- and posttests. GMRT results are provided as scale scores and as corresponding grade equivalent scores. For the purpose of our pre-post comparison, we used the grade equivalent scores because we needed a reference point that would tell us whether the students improved more or less than one grade level according to the norms used by the test. Grade equivalent measures consider the school year to be 10 months long, and each month is measured as 0.1 of a school year. In this case, because there were six months between test dates, that duration was considered 0.6 of a school year.

Results. The bar graph below shows the results of a *t* test comparing students' GMRT results from September 2004 and March 2005, in which the default assumption was that the posttest results should equal the pretest + 0.6. These students made improvement beyond what is expected for six months of schooling. Using the grade equivalent, we can interpret the difference of 0.244 as meaning the students made gains of about two and one-half months beyond expected gains.



Grade Equivalents Based on Fall 2004 and Spring 2005 Testing

A two-month advantage in the school context is significant from an educational point of view. The very low *p* value indicates that a difference this large or larger is very unlikely to have occurred by chance.

Conclusion. Our analysis of the data provides evidence for the effectiveness of *READ 180* in improving the reading scores of the district's 9th graders when the program is implemented adequately. We base our conclusions on the assumption that, without an intervention, students are unlikely to exceed the school-year growth expectations established for the reading test. It is, in fact, plausible to expect that high-school students struggling with reading will continue to fall behind. Comparing students' reading scores to growth expectations, we find that the *READ 180* students exceed the expectations by an educationally significant amount.

A larger and more rigorously controlled experiment is needed to determine whether an implementation that more closely follows the *READ 180* model would result in greater gains. With the limited implementation and without a control group to measure the actual achievement trajectory of struggling high school students who do not get an intervention, it is reasonable to conclude that this study may have underestimated the intervention's impact.

Table of Contents

Introduction

In the fall of 2003, Anaheim Union High School District began an implementation of Scholastic's *READ 180* program to help their middle- and high-school students who were identified as needing remedial reading. Scholastic contracted with Empirical Education Inc. in 2004 to work with the district in evaluating the program's impact. The district supplied us with data for the 9th graders for the 2004-2005 school year. Because the district had given all the students the Gates-MacGinnitie Reading Test, which provided both pre- and post-intervention measures, we were able to compare the reading achievement of students in classes taught using *READ 180* prior to and after the school year.

Our research question was whether students in classes using *READ 180* improve more than expected on measures of reading. Without a rigorously designed control condition, our study could not address causal relationships; numerous factors resulting from selection of students or teachers into *READ 180* could explain results in the kind of design used here. Nevertheless, with the goal of assisting the district in evaluating their initial implementation, our approach provides a useful metric.

Method

Research Design

The study is a pre- to post-intervention comparison of performance for students using *READ 180* during one school year. Although there was no control or comparison group, there was a normal expectation for one year of growth on the reading test. Therefore we were able to determine whether students made a significantly greater or lesser amount of growth than what was expected. If it is likely that 9th graders who are struggling readers may continue to lose ground each year, then this would be a conservative measure.

Materials

READ 180 is complex program with a specific implementation plan. Extensive documentation of the program and the existing research base is available on the Scholastic website (http://teacher.scholastic.com/products/read180/). The expectation is that reading classes are organized as 90-minute periods that are structured into a 20-minute whole group segment; 60 minutes of rotation through small group instruction, independent reading, and computer use; and a 10-minute wrap-up. The materials include interactive worktexts, paperbacks for independent reading, individualized instructional software, and audiobooks for modeled reading.

Site Description

Anaheim Union High School District serves more than 32,000 middle- and high-school students. The district is predominantly Hispanic (55%). White students make up about 25% of students and Asians are the next largest group (11%). Black students constitute 3.4% of the population. The district has organized reading classes for students who read below grade level and are not English language learners. Overall the district scores slightly higher than the California average in English Language Arts proficiency.

Sample

The district assigned students to reading interventions primarily on the basis of two reading scores: the California Achievement Test (CAT/6), a norm-referenced test which is part of the California testing

program, and the Gates-MacGinnitie Reading Test (GMRT). The students scoring below the 40th percentile on the CAT/6 or more than two levels below grade on the GMRT took a reading class that used *READ 180* or other interventions. Although the cut-off for *READ 180* was the 25th percentile, assignment was, in effect, determined by the availability of *READ 180* licenses. Thus the actual range for scores of students receiving *READ 180* differed somewhat from school to school, depending on availability.

The district reading specialist arranged for the student data to be provided to us for selected schools and teachers in which *READ 180* was adequately implemented. The sample consisted of 12 teachers and their classes. Table 1 shows the number of schools, teachers, classes, and students included in the data set.

Table 1: Sample broken down by number of schools, teachers, classes, and students

Level	Schools	Teachers	Classes	Students
Number of Units	6	11	39	645

While seven schools were initially included in the sample, one school was dropped because the implementation of *READ 180* in that school was hindered by technical and staffing difficulties. That school had four *READ 180* teachers.

Data Collection

Test Scores

The district provided us with student test score data. The pretest and outcome measures we intended to use were the student scores on the California Standards Test (CST), a standards-referenced test, for English Language Arts and the Gates-MacGinnitie Reading Test (http://www.riverpub.com/products/gmrt/details.html). CST Scores were obtained for end of year 2004 and 2005. We were unable to match this year's students with students exposed to *READ 180* in prior years. While there was a 12-month gap between the pre- and posttests for CST, the GMRT was given in September and March, so there was only a six-month difference between pre- and posttests. For the analysis reported here, the CST could not be used because a comparison of student change from one year to the next could not be made. The GMRT, however, provides a single growth scale as well as a grade equivalent; thus it was useful for the current analysis.

Survey of Teachers

Teachers of *READ 180* were given a web-based survey designed by Scholastic to measure aspects of program implementation. From this survey we derived four measures of differences that may impact the value of the program:

1) The number of semesters that the teacher had been implementing the program. We would expect teachers with more experience with the program to be better at taking advantage of all its features.

2) The number of hours that the teacher met with each class during the week.

3) The number of components of the program that were regularly used by the teacher.

4) How well equipped the classroom was with the materials and hardware needed for program implementation.

Although the small number of teachers responding to the survey does not allow an analysis of the impact of implementation differences on outcomes, it is useful to understand the nature and diversity of the usage.

Results

Implementation

The school year 2004-2005 was the district's second year of implementation for *READ 180*. Five of the 11 teachers had already used the program for a full year, two had begun half-way through the previous year, and for four, this was the first year of implementation. All the classes considered were at the 9th grade level (*READ 180* was in use in other grades as well) and all were using the high-school version of the product.

Although classes met each day, none of the schools organized the expected 90-minute periods. The district judged that changing the school schedules to accommodate the longer block would be very difficult, and teachers in the initial pilot had reported preferring the shorter periods. Consequently, most teachers had periods of 53 minutes, two teachers used 70-minute periods, and one teacher reported using a 40-minute period. Table 2 illustrates the range of implementations within the district with respect to the temporal elements that are part of the *READ 180* model. Only three of the teachers reported using the full sequence of whole-class instruction, rotation, and wrap-up. With the shorter periods, the element most likely to be dropped was the whole-class instruction at the beginning of each class. All the classrooms were well equipped with the books and other equipment that are an integral part of the program.

Table 2: Implementations of the *READ 180* classes in terms of length and structure of activities

Teacher	Length of class (in minutes)	Whole-group instruction at the beginning of each class	Rotation: small-group instruction, independent reading, computer use	Wrap-up
1	53	No	No	No
2	53	Yes	Yes	Yes
3	70	No	Yes	Yes
4	53	No	No	No
5	53	No	No	No
6	40	No	No	No
7	53	No	Yes	Yes
8	53	No	Yes	No
9	40	No	Yes	Yes
10	70	No	Yes	Yes
11	53	No	Yes	No

Ideally we would like to isolate the effects of specific aspects of implementation while controlling for the impact of all the others. However, because of the very small numbers of teachers, we did not attempt to draw any conclusions as to possible impact of these implementation differences on student achievement.

In most cases, students stay in the program for a whole year before either meeting criteria on one of the reading tests or simply moving on to the next grade. We did not have data that identified students who were in a class using *READ 180* during the prior year.

Attrition

We measured attrition of students by comparing the number of students taking the pretest to the number taking the posttest. Of the 645 students in the *READ 180* classes, 612 had pretest scores on the GMRT. Of these, 537 had scores on both pre- and posttests—a loss of 75 students. Thus student attrition was 12.3% between the two administrations of the GMRT. We also wanted to know whether the students who were lost during the year were different from those who remained. The lost students scored 6.87 points lower on the GMRT scale which, as a portion of a standard deviation, is 0.3. This difference should be considered in interpreting our findings, since they may not apply to the lowest scoring students.

Pre- to Post-implementation Gains on the GMRT

The GMRT results are provided as scale scores and as corresponding grade equivalent scores. For the purpose of the pre-post comparison here, we used the grade equivalent because we needed a reference point that would tell us whether the students improved more or less than one grade level according to the norms used by the test. Grade equivalent measures consider the school year to be 10 months long, and each month is measured as 0.1 of a school year. In this case, because there were six months between test dates, that duration was considered 0.6 of a school year.

Table 3 shows the results of a *t* test comparing student results from September 2004 and March 2005, in which the default assumption was that the posttest results should equal the pretest + 0.6. The results show that the students made improvement beyond what is expected for six months of schooling. Using the grade equivalent, we can interpret the difference of 0.244 as meaning the students made gains of about two and one-half months beyond expected gains.

Table 3: Dependent samples *t* test of the difference between posttest and adjusted pretest outcomes for GMRT Reading

Descriptive statistics for GMRT reading outcomes	Raw group means	Standard deviation	Number of students	Standard error
Spring 2005	7.170	2.122	537	0.0916
Fall 2004 (+0.6)	6.926	1.821	537	0.0786
***t* test for difference between means of dependent samples**	**Difference**		***t* value**	***p* value**
Posttest – adjusted pretest	0.244		3.94	<.0001

A two-month advantage in the school context is significant from an educational point of view. The very low *p* value indicates that a difference with an absolute value this large or larger is very unlikely to have occurred by chance. This difference is shown graphically in Figure 1.

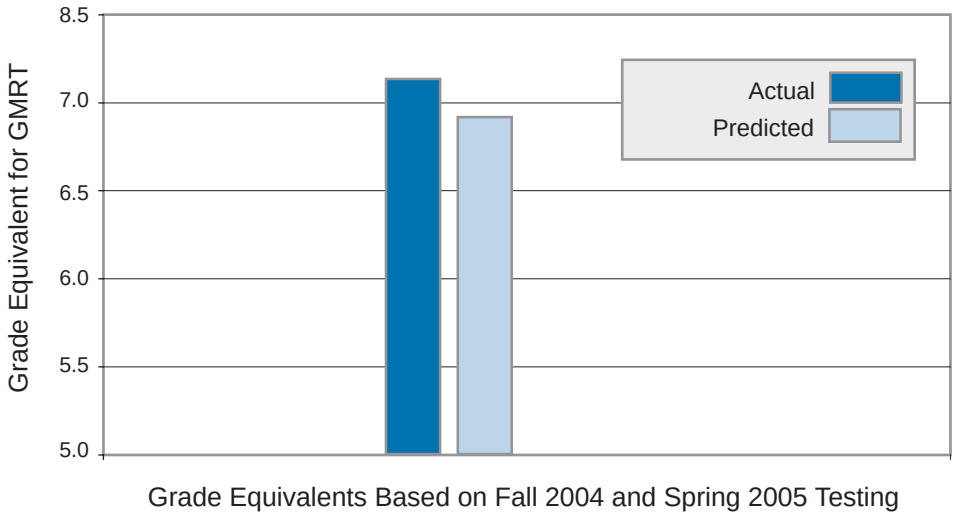


Figure 1: Comparison of predicted and actual GMRT grade equivalent scores for the students in *READ 180*

Discussion

Our analysis of the data provides evidence for the effectiveness of *READ 180* in improving the reading scores of the district's 9th graders when the program is implemented adequately. We base our conclusions on the assumption that, without an intervention, students are unlikely to exceed the school-year growth expectations established for the test. It is, in fact, plausible to expect that high-school students struggling with reading will continue to fall behind. Comparing students' reading scores to growth expectations, we find that the *READ 180* students exceed the expectations by an educationally significant amount. Using a statistical test, we find that a difference that large is very unlikely to have occurred by chance.

In this pre- to post-intervention analysis of performance for students in the *READ 180* program, the research design did not include a matched comparison group. If it is a reasonable assumption that students who are struggling readers will continue to fall behind, then a study using a well matched, or randomly assigned, control group design might have demonstrated an even larger effect.

Our sample of only 11 teachers does not allow an analysis of the impact of implementation differences. None of the *READ 180* classes followed exactly the implementation plan that is called for by the program. A larger and more rigorously controlled experiment is needed to determine whether an implementation that more closely follows the *READ 180* model would result in greater gains. With the limited implementation and without a control group to measure the actual achievement trajectory of struggling high school students who do not get an intervention, it is reasonable to conclude that this study may have underestimated the intervention's impact. Within the limits of the current study, the research supports the district's decision to implement *READ 180*.