RESEARCH REPORT

Effectiveness of Scholastic's READ 180 as a Remedial Reading Program for Ninth Graders:

Report of a Comparison Group Study in Poway, CA

Denis Newman
Empirical Education Inc.

Minh-Thien Vu
Empirical Education Inc.

May 1, 2006

Empirical Education Inc.
www.empiricaleducation.com
425 Sherman Avenue, Suite 220
Palo Alto, CA 94306
(650) 328-1734

Acknowledgement

We are grateful to the Poway Unified School District for providing the data for this analysis and to Scholastic, Inc., which sponsored the research.

About Empirical Education Inc.

Empirical Education Inc. was founded to help school districts, publishers, and the R&D community assess new or proposed instructional and professional development 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., & Vu, M. (2006, May). Effectiveness of Scholastic's READ 180 as a remedial reading program for ninth graders: Report of a comparison group study in Poway, 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 Poway Unified School District to evaluate the impact of its READ 180 program in improving reading performance for 9th graders during the 2004-2005 school year. The study sought to determine whether students in classes using READ 180 improved more than expected on measures of reading, whether they outperformed their peers in regular Reading classes, and whether creating special classes for READ 180 provides better results than the regular Reading classes. The study's findings do not provide evidence of a beneficial effect of READ 180.

Intervention. READ 180 is complex program with a specific implementation plan; classes are organized as 90-minute periods segmented as whole- and small-group instruction. Materials include interactive worktexts, paperbacks for independent reading, individualized instructional software, and audiobooks for modeled reading. The regular 9th grade Reading program consisted of a set of readings from several books rather than a single publisher's program.

Setting. Poway Unified School District, located in a relatively affluent suburban community near San Diego, serves about 33,000 students K–12, about 75% proficient in English Language Arts—considerably above the 45% California overall average. The district population is 61.7% White, 13.8% Asian, and 9.6%Hispanic. Our study was based on data from the four of the district's five high schools, all of which are similar in English proficiency levels and demographics.

Research design. The district supplied pre- and post-intervention scores on the California Standards Test (CST) for English Language Arts. We compared outcomes for 9th graders taught using READ 180 and 9th graders with similar pre-intervention scores who participated in the district's regular Reading program. In an analysis of covariance (ANCOVA), we used the pretest score to control for initial group differences. A multilevel mixed statistical model provided a measure of these differences and an indication of their significance. For a subset of READ 180 students, we also used the NWEA MAP reading test to examine growth over a school year.

Participants. READ 180 is used in special classes for struggling readers in two Poway USD high schools, where six teachers provide READ 180 instruction for one class apiece. Comparison group students were chosen from among those served in regular Reading classes at the four participating high schools on the basis of similar pre-intervention test scores. Comparison group students had higher initial scores than READ 180 students.

Implementation. Three of the four READ 180 teachers were in their first year to use the program in their teaching. Although all had received training, three of the four met with their classes for fewer minutes per day than stipulated in the READ 180 implementation model; one teacher, whose class time was substantially below expectation was dropped from the study. Limited computer availability may have adversely affected use of the READ 180 software.

Statistical Analysis. SAS PROC MIXED, our primary tool for the initial between-group adjustment noted above, allowed us to account for the clustering of students in classes (intra-class correlation) and provided an accurate assessment of the confidence we should have in the findings. We also used a t test to compare the mean of the predicted score based on norms to the actual outcome for READ 180 students who had taken the NWEA MAP reading test.

Results. Outcomes for between-group comparisons were for the English Language Arts section of the CST. Raw means for the comparison group, whose initial scores were higher, surpassed the mean scores for the READ 180 group—even when the analysis focused on the three teachers whose implementation was consistent with the model. Although adjusting for the initial difference in scores resulted in a small positive value for READ 180, the high p value suggests that this difference could be a result of chance.

We found similar results for a within-group analysis using NWEA MAP test scores for students in the three classes where READ 180 was adequately implemented. Based on pretest scores and using the growth norms established by NWEA, we calculated the expected growth for each student. When we compared the normal expectations with the actual growth for the year, we found a small positive difference that could have occurred by chance. Still, in the education context, these findings may have practical value. If we assume that students would have lost ground without an intervention, then it could be argued that the gain is underestimated.

Conclusion. This comparison did not find evidence that achievement for students in READ 180 achieved better in reading than they would have done in regular Reading classes. Neither our comparison to similar non-READ 180 students nor our comparison of READ 180 students to growth norms revealed a difference that reached a reasonable level of statistical confidence. A finding of no difference does not mean that READ 180 is ineffective. Information gained from teacher surveys suggests that start-up problems or lack of familiarity with the program may have reduced its impact. We recommend that the district continue monitoring effectiveness of the program. Improved implementation through better access to computers and greater familiarity with READ 180 on the part of teachers may contribute to improved reading achievement. More systematic use of formative testing such as provided by NWEA will provide a more precise measure of differences in growth on which to base decisions.

Effectiveness of Scholastic's READ 180 as a Remedial Reading Program for Ninth Graders:

Report of a Comparison Group Study in Poway, CA

Table of Contents

Introduction

Administrators at Poway Unified School District were interested in evaluating the impact of Scholastic's READ 180 program that was in use in two of their four high schools for students identified as needing remedial reading instruction. The district provided data for the 2004-2005 school year. We were able to compare the English Language Arts achievement on the California Standards Test of students in classes taught using READ 180 to that of comparable students attending the district's regular Reading classes. For a subset of students in READ 180, we were also able to compare the pre- and post-intervention scores on a reading test that provided a metric with norms for the expected growth over a school year.

This study was intended to provide useful information to support local decisions in Poway USD but not to generate broadly generalizable results. The results should not be considered to apply to school districts with practices and populations different from those found in Poway.

Method

Research Design

This research study is a comparison of outcomes for students taught using the READ 180 program and those taught using the district's regular Reading program. The district supplied test scores for all 9th grade students in the READ 180 program as well as for 9th graders whose scores were in the same general range. Most students in the sample had pre-intervention scores on the California Standards Test for English Language Arts from the previous spring (2004) and post-intervention scores from the following spring (2005). In an analysis of covariance, we used the 2004 score to control for initial differences between the groups. A multilevel mixed statistical model provides a measure of the difference between the groups of students and an indication of its significance.

The two conditions were different in that READ 180 was used in the context of a scheduled Reading class specific to that program. Because students selected as comparison group members were immersed in regular 9th grade Reading classes, the number of those students per class was small (often only one). Realistically, this is the comparison that the district needs to understand: whether creating special classes for READ 180 provides better results than leaving the students in the regular Reading classes.

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

Poway Unified School District serves about 33,000 students in Kindergarten through grade 12. The district's percentage of students who are proficient in English Language Arts—about 75%—is

considerably higher than the percentage for California as a whole—about 45%. The predominant ethnicity is White at 61.7%, followed by Asian at 13.8% and Hispanic at 9.6%. The district is located in a suburban community 15 miles northeast of San Diego. It is primarily a middle- to upper-middle-income residential community.

The district has five high schools, including one continuation school. Our study was based on data from the four regular comprehensive high schools. Proficiency levels and demographics among these four were very similar.

The regular Reading program for 9th graders consisted of a set of readings taken from several books. It was not a textbook program from a single publisher.

Sample

In Poway USD, READ 180 is in use in two of the high schools, where six teachers each provide READ 180 instruction for one class that is specially composed of struggling readers. Other readers in the district are served in regular Reading classes. Comparison group students were chosen from among these other students at all four of the high schools on the basis of their pre-intervention test scores being similar to READ 180 students. Two of the READ 180 teachers also taught Reading classes containing some students eligible for the comparison group.

Data Collection

Test Scores

We obtained scores for the spring 2004 and spring 2005 administration of the English Language Arts subtest of the California Standards Test (CST), a criterion-referenced test which is tied to the California standards in specific content areas. The CST is broader in scope than just reading ability and is possibly a less sensitive test for reading than for other components of English Language Arts. It is not possible to compare CST scores from year to year as a growth measure because the test for each grade addresses a different set of standards. However, we can use the previous year's scores statistically to adjust for initial group differences.

For some students in the sample, we also had the Northwest Evaluation Association's MAP reading test results for spring 2004 and 2005. Although the MAP test was not given consistently throughout the district, we have NWEA results for most students in three of the READ 180 classes. The MAP test is a computer-adaptive test of reading, meaning that the questions are made easier or harder dynamically as students progress through the test. It results in a continuous growth scale that has been tested with populations of students in order to measure normal growth during a year of schooling.

Survey of Teachers

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

1) the length of time 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 minutes 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 in the study does not allow an analysis of the impact of implementation differences on outcomes, the responses are useful for understanding the nature and diversity of the usage.

Statistical Analysis

An analysis of covariance (ANCOVA) allows us to control for differences among students in their initial achievement level. We used SAS PROC MIXED (from SAS Institute, Inc.) as the primary tool for this work. This software is particularly appropriate in research on schools where the outcomes are measures of individual student performance because it allows us to account for the clustering of students in classes (the intra-class correlation) and provides a more accurate, and often more conservative, assessment of the confidence we should have in the findings. We also used the simpler t test to compare the mean of the predicted score based on norms to the actual outcome for READ 180 students who had taken the NWEA MAP reading test.

Results

Implementation

Our information about implementation is based on the teacher survey described above. For all the READ 180 teachers except one, this was their first year to use the program in their teaching. All had received training.

Table 1: Implementation for READ 180 classes in terms of length and structure of activities

Teacher	Class minutes per week	Whole-group instruction at the beginning of each class	Rotation: small-group instruction, reading, computer use	Wrap-up	Sufficient computers available
1	60	No	Yes	No	No
2	270	Yes	Yes	No	Yes
3	504	Yes	Yes	Yes	Yes
4	504	Yes	Yes	Yes	Yes
5	512	Yes	Yes	Yes	No
6	374	Yes	No	Yes	No

Table 1 indicates that three of the teachers met with their classes for fewer than the 90 minutes per day stipulated in the READ 180 implementation model. In one case, the amount of time spent was substantially below the expectation. The introduction, rotation, and wrap-up were carried out consistently

by half the teachers. In addition, because computer availability was less than optimal for half the teachers, rotation through the READ 180 software may not have been consistent.

Teacher 1 reported providing three 20-minute periods per week. In the survey, this teacher reported: "Given the immensity of technical breakdowns we've experienced all year, it's hard to tell how effective this program really is." In view of this discrepancy, we chose to drop this class from the subsequent analysis.

Initial Group Composition

The samples included students registered in classes that used READ 180 and other students from the district with a similar range of test scores who were enrolled in regular 9th grade Reading classes. Table 1 shows the initial distributions of students' initial CST scores on the English Language Arts subtest.

Table 2a: Comparison of pre-intervention CST scores in READ 180 and comparison groups

Group	Mean	Standard deviation	Number of students
READ 180	290.822	34.983	73
Comparison	299.607	35.364	117

It is apparent that the comparison group students have a higher initial mean achievement level than students in the READ 180 group. The differences in starting point for these two groups were adjusted statistically in the model used for comparison of the groups reported later.

Table 2 provides a count of the teachers, classes, and students in the sample. The comparison group includes a far greater number of classes because the students with low test scores were distributed among regular Reading classes.

Table 2b: Count of teachers, classes, and students in READ 180 and Comparison groups

READ 180 vs. Comparison			
Condition	Teachers	Classes	Students
READ 180	5	5	86
Comparison	32	57	147
Totals	34	62	204

*Three teachers taught both READ 180 and regular Reading classes.

Attrition

Of 190 students who took the pre-intervention test, 24 students did not take the posttest. This yields an attrition rate of 6.3%.

Quantitative Comparison of Outcomes

Comparison of CST Outcomes between READ 180 and Comparison Groups

Our first step in the analysis of between-group differences was to compare the outcomes for the classes of the five teachers who reported using READ 180 a significant amount of time with the outcomes for students in the regular 9th grade Reading classes. As noted, outcomes for this study were for the English Language Arts section of the CST. Our statistical model is reported in Table 4. We used a multi-level mixed model; that is, we used the clustering of students in classes as part of the model. The table shows that the raw means

Table 4: Multi-level mixed model for CST ELA—results controlling for pretest scores

Descriptive statistics: CST ELA outcomes	RAW Group Means	Standard Deviation	Number of Students	Number of Classes	
Read 180	306.275	37.806	69	5	
Comparison	309.020	37.510	102	57	
Mixed model: Fixed factors related to CST ELA outcomes	**Estimate of Coefficient**	**Standard Error**	**DF**	**t value**	**p value**
Interest	305.11	2.837	31	107.56	<.0001
Pretest score (centered at the mean)	0.771	0.066	125	11.74	<.0001
Condition (READ 180 = 1; comparison = 0)	2.321	5.162	125	0.45	0.654
Mixed model: Technical details for random components	**Estimate of Variance Component**	**Standard Error**		**z value**	**p value**
Class mean achievement	15.334	68.741		0.22	0.412
Within class variation	716.73	89.269		8.03	<.0001

Note: This model is based on 204 cases. Missing cases include 33 students without posttest scores. Another 12 cases were removed as outliers or influential points.

for the comparison group were slightly higher than mean scores for the READ 180 group. This was expected because, as Table 1 shows, the comparison group started out ahead. The statistical model adjusts for this initial difference and results in a small positive value for READ 180 (2.321 points on the CST scale). However, the p value for this is very high, indicating that it is reasonably likely that a difference of this size could be a result of chance.

Since there was variation in the implementation shown in Table 3, we also ran an analysis comparing the results for the three teachers who spent more than 450 minutes per week (consistent with the READ 180 implementation model). Although the results for these teachers showed a smaller difference from the comparison group, again the p value was very high. With a small number of teachers, it is possible that other factors such as teaching style may have played a role.

Comparison of READ 180 between Predicted and Actual Reading Scores

Although we had data for the NWEA MAP reading test for only 57 of the READ 180 students, we decided

to do an analysis of this set drawn from three classes that implemented READ 180 adequately and in which most students had NWEA test scores. The analysis in this case is not a comparison to the non-READ 180 students. Instead, based on the pretest and using the growth norms established by NWEA, we calculated the expected growth for each of the READ 180 students. We then compared the normal expectations with the actual growth for the year. Table 5 reports a t test for the difference between the observed and predicted outcomes.

Table 5: t test of the difference between observed outcomes and predicted outcomes on the NWEA reading test

Descriptive statistics: NWEA score outcomes	Raw Group Means	Standard Deviation	Number of Students	Standard Error	Effect size
Observed score Predicted score	217.37 215.79	11.41 9.16	57 57	1.511 1.214	0.153
t test for difference between dependent means	Difference		DF	t value	t value
Observed and predicted score	1.58		56	1.21	0.2310

There was a small positive difference, but the p value of .231 suggests a good likelihood that the difference could have occurred by chance. Although an effect size of .153 is considered small, in the education context, this finding may have practical value. If we assume that students would have lost ground without an intervention, then it could be argued that the gain is underestimated.

Discussion

This study does not provide evidence of a beneficial effect of READ 180. Neither our between-group comparison to similar students not in READ 180 classes nor our within-group comparison of READ 180 students to growth norms revealed a difference that reached a reasonable level of statistical confidence. A finding of no difference does not mean that the program is ineffective. It simply means that, in this comparison, it was in the same range of effectiveness as the programs currently in place.

Our cursory survey of the teachers' usage indicated that there were some deviations from the model for READ 180 implementation recommended by Scholastic. For two of the teachers whose classes were included in the analyses, the amount of time devoted to READ 180 activities was lower than expected, and the classroom activity structures did not completely match the expectation. In addition, half the teachers reported inadequate access to computers. Moreover, while all teachers reported that they received training, for all but one of them, this was their first year to use the program. Start-up problems or lack of familiarity with the program may have reduced its impact. Research methodology can contribute to inaccuracies and biases. Random assignment of eligible students into READ 180 classes or regular Reading classes will remove any bias associated with informal methods of selecting students for the program. A comparison of the achievement of two groups of students created by random assignment and thus known to be equivalent at the study's outset would provide much greater protection

against bias than is possible with pre-selected groups. For example, the uncertainty as to how much the READ 180 students would have grown without the intervention would be addressed in a randomized experiment. More detailed teacher surveys and observations of implementation will also help to identify implementation difficulties that can interfere with program effectiveness.

This comparison did not find evidence that students in READ 180 achieved better in reading than they would have done in regular Reading classes. We recommend that the district continue monitoring effectiveness of the READ 180 program. Improved implementation through better access to computers and greater familiarity with the program on the part of teachers may contribute to improved reading achievement. More systematic use of formative testing such as provided by NWEA will provide a more precise measure of differences in growth on which to base decisions.