Empirical Education

RESEARCH REPORT

The Efficacy of *PCI's Reading Program - Level One*:

A Report of a Randomized Experiment in Brevard Public Schools and Miami-Dade County Public Schools

Megan Toby
Boya Ma
Andrew Jaciw
Jessica Cabalo
Empirical Education Inc.

October 2008

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

Acknowledgements

We are grateful to the people in Brevard Public Schools and Miami-Dade County Public Schools for their assistance and cooperation in conducting this research. The research was sponsored by PCI Education, which provided Empirical Education Inc. with independence in reporting the results.

About Empirical Education Inc.

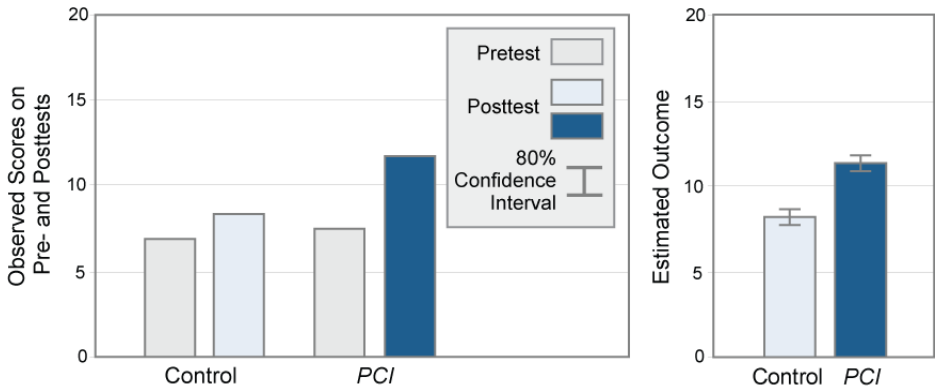
Empirical Education Inc. was founded to help school districts, publishers, and the R&D community assess new or proposed instructional and professional development programs through scientifically based pilot implementations. The company draws on the expertise of world-class researchers and methodologists 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.

Executive Summary

Introduction. PCI Education sought scientifically based evidence on the effectiveness of the *PCI Reading Program - Level One* for students with severe disabilities. During the 2007-2008 academic year, Empirical Education conducted a randomized control trial (RCT) in two Florida districts, Brevard and Miami-Dade County Public Schools. For this experimental study, we randomly assigned teachers to two groups: a group trained on and using the *PCI Reading Program - Level One* (*PCI* group) and a control group using their existing reading program. The specific question we addressed is whether students whose teachers were given the *PCI Reading Program - Level One* achieve higher sight word assessment scores in reading than students of teachers not having it. We were also interested in whether effects differed for specific subgroups of students: those who score low on the sight word or phonological pretests, those in lower or higher grades, and those whose teachers were more experienced in Special Education. PCI Education also asked us to investigate the level of implementation that could be achieved and the level of interest generated among the teachers. As an initial study of this program (Levels Two and Three are still in development) our goal was to determine whether or not it helps severely disabled students succeed in learning the specific skills on which it is focused. Because our outcome measure, sight word recognition, was closely aligned to the program (Level One is focused on sight words) we consider this an efficacy study in which we are examining whether the program achieves its stated goals rather than whether it makes a difference in more generally defined reading skills.

Findings. Despite attrition among students and teachers, the experiment was able to detect a very large impact: After one year, students in the *PCI* program had substantial success in learning sight words in comparison to students in the control group—equivalent to a 21 percentile point difference.

We accommodated students unable to name any words on the pretest by conducting separate tests for those scoring zero on the pretest and those scoring above zero, as we believed the composition of these groups could differ fundamentally. Within each group, we found a significant impact for the *PCI* program. Both unadjusted and adjusted analyses show high effect sizes (.55, .59) with small *p* values.



For Sight Word Assessment: Unadjusted Pre- and Posttest Means for Control and *PCI* (Left); Adjusted Means for Control and *PCI* (Right)

In examining moderating variables, we found the sight word pretest to not be significant in changing the impact of *PCI* on student outcomes. The Phonological Assessment had a small moderating effect; thus we can have some confidence that students starting with greater phonological skills benefit more from *PCI* than students scoring lower. We found no effect from the student's grade level. The small number of teachers with fewer than four years experience teaching Special Education prevented us from examining moderating effects of teacher experience.

Although teachers in the *PCI* group had all materials necessary for implementation, they tended to complete the minimum amount of reading instruction specified by the publisher. Because we do not know how much time *PCI* instruction was supplemented with other reading materials, we were unable to compare minutes of reading instruction across the control and *PCI* groups. Still, our study

shows that it takes longer for students to complete the *PCI Reading Program - Level One* than initially estimated by the publishers.

Overall Teacher Impressions. Qualitative data obtained from surveys, observations, and informal interviews showed very high levels of teacher satisfaction, as well as student engagement and enjoyment, with *PCI*. Teachers assigned to the *PCI* group for the purpose of this study plan to continue implementing the program in the future. While some teachers reported difficulty in finding time for the amount of individualized instruction necessary, opinions of the program itself were high. Teachers were excited to see their students retaining words they learned through the program and reading books.

Design and analysis. The design of our experiment was a group randomized trial. We used a coin toss to assign teachers in two Florida school districts (Brevard and Miami Dade County Schools) to use the *PCI* program or to continue using the various reading programs currently in place. We used paired randomization and a coin toss, to assign each of the 47 teachers initially involved in the experiment to the *PCI* and control conditions. The teachers gave a sight word test based on words used in the *PCI* program and words common to other programs both pre and post. Significant attrition of students occurred because of difficulty in obtaining parental consent. Information on implementation was gathered through online surveys, observations and teacher interviews. Multi-level analysis of covariance (hierarchical linear modeling) was used to estimate program impact and the moderating effect of relevant variables.

This initial experimental study provides evidence of the efficacy of the *PCI Reading Program*. The positive results for students and positive acceptance by the teachers is useful information for school districts looking for a reading program for severely disabled students. We consider our results preliminary because we tested only *Level One* and our achievement measure was limited to the specific goals of the program. We are continuing our research of the *PCI Reading Program* in both districts over the next four years in order to follow students through the second and third levels of *PCI*.

The Efficacy of PCI's *Reading Program - Level One*:

A Report of a Randomized Experiment in Brevard Public Schools and Miami-Dade County Public Schools

Table of Contents

Introduction

PCI Education contracted with Empirical Education Inc. to conduct a one-year randomized control trial (RCT) to determine the efficacy of the *PCI Reading Program - Level One (PCI)* as implemented in two sites. We report here on the research that began in April 2007 in the Brevard Public Schools (BPS) and Miami-Dade County Public Schools (MDCPS) on the program's efficacy for reading achievement among students with severe disabilities.

The specific question we addressed is whether students whose teachers have been given the *PCI Reading Program - Level One* achieve higher sight word assessment scores in reading than students whose teachers do not have the program. We were also interested in whether *PCI* had a different effect for specific subgroups of students: those who score low on the sight word or phonological pretests, those in lower or higher grades, and those who had teachers with more experience teaching special education. PCI Education also asked us to investigate the level of implementation that could be achieved and the level of interest generated among the teachers. Because this was an efficacy study, the program's effectiveness was not tested against standardized measures of reading. The outcome of interest was a sight word test aligned with the *Level One* program being tested. As an initial study of this program, (Levels Two and Three are still in development), our goal was to determine whether or not it helps severely disabled students succeed in learning the specific skills on which it is focused.

For this experimental study, we randomly assigned 47 teachers into two groups: a group of teachers who were trained on and used the *PCI Reading Program - Level One* (*PCI* group) and a control group that continued with their existing program, that is, with "business as usual." A randomized experiment or randomized control trial eliminates the variety of biases that could otherwise compromise the validity of the research. For example, it ensures that teachers in both groups are equivalent in their interest in trying *PCI* and in their ability to take advantage of the new program.

Random assignment of subjects to experimental conditions does not assure that we can generalize the results beyond the districts where the research was conducted. We designed our study to provide useful information to support local decisions that take into account the specifics of district characteristics and district teachers' implementation of the program. The results should not be considered to apply to school districts with practices and populations different from those in this experiment. The report presents a description of the conditions of initial program implementation and provides the reader with an understanding of the context of both district sites.

Methods

Our experiment is a comparison of outcomes for teachers where *PCI* was in place and for a control group of teachers using their districts' current methods. This section details the methods used to assess, with some level of confidence, the size of the difference in outcomes and whether the introduction of *PCI* was responsible for those differences. We begin with a description and rationale for the experimental design and then describe the intervention, the research sites, the sources of data, the composition of the experimental groups and, finally, the statistical methods used to generate our conclusions about the impact of *PCI*.

Experimental Design

Before beginning the experiment we established the specific questions to be answered, including identifying the important outcomes and the population to which we want our conclusions to apply. The upfront investment of time to fully specify a design or plan for the experiment pays off by focusing our effort on clearly defined questions. It also avoids "fishing" for results after the fact that could lead us to conclude that an effect exists when it is just a chance occurrence. Part of the essential preliminary work is to determine how large an experiment should be in terms of students, teachers, and schools in order to reach the desired level of confidence in the results.

Research Questions

We identified up front the following questions, which pertain to students and teachers to whom we can generalize our results:

1. Do students of teachers who use *PCI Reading* learn to recognize sight words more successfully than students of teachers using the existing district reading programs for that population?
2. Do students who score lower on the sight word or phonological pretests benefit more from *PCI* than students who score higher on the pretests?
3. Do students in lower grades benefit more from *PCI* than those in higher grades?
4. Do students of teachers with four or more years of special education teaching experience benefit more from *PCI* than those who have teachers with fewer than four years experience teaching special education?

In addition to these experimental questions we also planned to document the implementation of the program.

Randomization

Because we seek to know the impact of *PCI*, if any, we must isolate its impact from all other factors that might make a difference for how or what teachers and students do. Our goal is to answer definitively whether *PCI* caused any differences we might observe. Randomization ensures that, on average, characteristics other than the intervention that affect the outcome are independently distributed between program and control groups. This distribution prevents us from confusing the intervention's effects with some other factors—technically called "confounders"—that are not randomly distributed between groups and that can affect the outcome. For example, randomization helps to ensure that more experienced teachers are not selectively assigned to the program group or to the control group.

There are various ways to randomize teachers to experimental conditions. For this study, we used a matched-pairs design where we identified pairs of similar teachers. First, we consider what the critical characteristics of teachers are that we believe affect performance. We use this information to pair teachers, as far as is possible, and then we randomize the members in each pair to the two conditions—experimental and control. Technically, this is a form of blocking that usually increases the precision of our impact estimate. As previously noted, this experiment's matches were based on grade level taught, current reading program, number of students in the class, and years teaching special education. We used a random process – a coin toss – to assign teachers to conditions.

Organizational Levels Considered in the Experiment

This research study works within the organization of the participating schools by not disrupting the existing hierarchy in which students are grouped under teachers in the schools. The level in the hierarchy at which we conduct the randomization is generally determined on the basis of the kind of intervention being tested. School-wide reforms call for a school-level randomization, whereas a professional development program can use a teacher-level randomization. Generally, we attempt to identify the lowest level at which an intervention can be implemented without unduly disrupting normal collaboration and without inviting sharing or "contamination" between control and program units. For this experiment, we randomized teachers who volunteered for participation in approximately equal numbers to the *PCI* and control groups. The outcome measures are student-level test scores on the Sight Word Post-Assessment that was developed specifically for this study. Because teachers rather than students were assigned to *PCI* or control groups, this kind of experiment is often called a "group randomized trial."

What Factors May Moderate the Impact of *PCI*?

Our research design allows us to consider the extent to which *PCI* is differentially effective for students at various points along the pre-assessment scale (in both the Sight Word Pre-Assessment

and Phonological Pre-Assessment) as well as for students at different grade levels and for students whose teachers have more than four years of special education teaching experience. These variables are measured before the experiment starts, as we have reason to believe that they will influence the strength of the effect of the *PCI Reading Program*. Technically, variables such as these are called potential moderators because they may moderate the impact of the treatment. During analysis we measure the strength of the interaction between each moderator and the *PCI Reading Program* effect; that is, we measure whether the effect of *PCI* changes as the level of the moderator changes.

How Large a Sample Do We Need?

A process called power analysis was used to plan the number of teachers that the experiment needs in order to say with any confidence that the intervention has an impact of a certain size. This is an important part of experimental design and here we walk through the factors considered.

How Small an Impact Do We Need?

The size of the sample required for a study depends on how small an effect we need to detect. Experiments require a larger sample to detect a smaller impact. It is very important to make an educated guess about the range of impact typical for an intervention similar to the one being tested. On a practical level it is also important to know the smallest potential impact that would be considered educationally useful in the study's particular setting. As a hypothetical example, using percentile ranks as the measure of impact, we may predict that an intervention of this type can often move an average student 15 percentile points. As a practical matter for educators, however, an improvement as small as 10 percentile points may have value. The researcher may then set the smallest effect of interest to be 10 points or better. Thus if the intervention makes less than a 10-point difference, the practical value will be no different from zero. Such a pre-determined effect size level is referred to as the "minimum detectable effect size" (MDES) for the experiment. It is necessary to decide in advance on this value as part of the power analysis. In some cases, positive effects may exist that we cannot detect because they are lower than the MDES. For the current experiment with *PCI*, the design and sample size were adequate for an MDES of 14 percentile points or, in terms of the standard deviation units we introduce below, for an effect size of .35.

How Much Variation Exists Between Teachers?

When we randomize at the teacher level with the outcome of interest being a test score of students associated with those teachers, we must pay special attention to the differences among teachers. The greater the differences among them, the more teachers we need in the experiment to detect the impact of the intervention. This is because the extra variation among teachers adds "noise" to our measurement which makes the effect of the intervention—the "signal"—harder to detect. A larger sample allows us to effectively reduce the level of the noise. If the differences among the teachers, including the teacher averages of student characteristics, are large, and/or the differences among students within teachers are small, then the sample size that matters the most for the experiment is the number of teachers. If the differences among teachers are very small, then the sample size of students begins to matter more. A summary statistic, the intraclass correlation (ICC), tells us how the variation is divided up among levels of analysis. Technically the ICC is the ratio of the variation in the outcome among teachers to the total variation in outcome. To perform the power analysis we have to assume a plausible value for the ICC. For this experiment we assumed a fairly conservative intraclass correlation of .20.

How Much Value Do We Gain From a Pre-Assessment?

In order to gain additional precision, we make use of other variables that we know will impact performance. In our experiments, a student's score on a pre-assessment (which may be a test in a subject that is closely related to the outcome measure rather than the same test given earlier) is almost always the variable most closely associated with the outcome. In this case, the pre-assessment is a "covariate." By including the covariate in the analysis we can increase

precision by "removing" this source of variation in the results. Technically, a covariate-adjusted analysis is called an analysis of covariance (or ANCOVA). In nearly all our analyses, we adjust for the effect of the pre-assessment, which is a strong predictor of posttest performance. In this experiment, we assumed a fairly substantial correlation between the pre- and posttests (.80[1]). In a power analysis determining the number of teachers we will need, a good pre-assessment correlation will increase precision and thereby require fewer teachers to detect the same level of impact.

Are There Subgroups of Particular Interest?

Often we are interested in whether a program has more impact for a particular student subgroup than others or for a certain group of teachers than others. Where the subgroup is identified within each randomized unit—that is, where each randomized unit has some portion of that subgroup—there is no adverse effect on power. However, if our subgroup of interest is a subtype of the unit of randomization, then, in most cases, we would need to include additional units in the experiment in order to detect differential effect of the intervention with as much power as we have to detect the average effect. Alternatively, we acknowledge that we have less power to detect such moderating effects, and consider such analyses exploratory. In the current experiment, we are interested in differences between students with low prior achievement on either the sight word pretest or a test of phonological skills as compared to the rest of the students, students at different grade levels, and teachers with fewer than 4 years of teaching experience in special education as compared to teachers with more experience.

How Much Confidence Do We Want to Have in our Results?

We have described uncertainty in terms of the likelihood that, if we ran the experiment again with a different sample from the same district, we would get the same result. Although results are never exactly identical, we can design the experiment so that the various results we would get would be similar. This scenario is hypothetical because we are not likely to run exactly the same experiment multiple times. An experiment that produces a very high level of confidence that the results of multiple experiments would be very similar requires a larger number of units than an experiment that produces a lower level of confidence or a wider range of likely outcomes for the other hypothetical experiments. Still, we can never be entirely certain of a result. Thus the final step in the power analysis is to determine an acceptable or tolerable level of uncertainty. Conventionally, researchers have called for a high level of certainty, specifically, that getting a result as large or larger than the one observed would happen in only 5% of instances if the program did not indeed have an impact. For the purpose of the power analysis for this experiment, we used the 5% criterion although, as we explain later, we report the results using a range of confidence levels.

Sample Size Calculation for This Experiment

Taking all the above factors into consideration, we estimated that 50 teachers (twenty five in each experimental group) would constitute a sufficiently large sample to detect an effect size as small as .35. (As we describe later, due to several factors, the sample size of teachers that we ended up using in the analysis was smaller than this, but we detected a strong impact nonetheless.)

---

[1] That is, we assume that the square of the correlation (i.e., .80*.80=.64) is the proportion of variance in the outcome (i.e., the R-squared) that is accounted for by the covariate, in either condition.

Intervention

The intervention we are testing consists of the *PCI Reading Program - Level One* kit and a one-day training for the teachers.

Training/Professional Development

Teachers from both sites were invited to participate in a one-day training to familiarize participants with the *PCI Reading Program - Level One* and the research study. Trainings for the two districts occurred at separate times and locations. Both *PCI* and control group teachers attended the training for an hour-long session to introduce the study and its expectations, as well as to train the teachers to administer the pre-assessment. Members of the control group were not present for the remainder of the day, during which *PCI* group teachers learned about implementing the *PCI Reading Program - Level One*.

The *Level One* training, for *PCI* teachers only, was led by Jill Haney of *PCI* Education. Jill Haney is one of the authors of the program, a former classroom teacher, and the POC from *PCI* Education. During the training Haney introduced teachers to the literature and rationale behind *PCI*. In addition, Haney commented on the need for research within the student population targeted by this reading intervention program. The training began with a theoretical and pedagogical overview of *PCI*. A brief introductory video of the program was followed by a question-and-answer session. After learning about what teachers and students could expect from the program, each teacher received a *PCI Reading Program - Level One* kit. Haney spent the remaining hours familiarizing teachers with the materials and leading them through the steps in the lesson cycle. She demonstrated lessons and showed teachers how to institute practice exercises so they would understand how the materials and the program would work within the context of their own classrooms. Haney stated that the program was intended to be followed quite directly, with a high degree of fidelity to the Teacher's Guide. It was made clear that there would be no follow-up training, but that participating teachers could feel free to contact *PCI* Education or Empirical Education Inc. with any subsequent questions.

All participating BPS teachers attended the full-day training session. All participating MDCPS teachers either attended the initial full-day training or received make-up training. One MDCPS teacher was unable to attend the training and had to drop out of the study due to a family emergency. Another MDCPS teacher, who had not yet been randomized, attended the pre-assessment training and wanted to join the study. This teacher was randomized after the training, along with three other teachers who joined the study post-training. Also present at the MDCPS training were the district POC and two assistants. The two assistants attended so that they could administer make-up trainings to teachers who were unable to attend.

Make-up trainings were held in MDCPS on an individual basis during the month following the initial training. A total of seven teachers received make-up training (only pre-assessment training for the control group; complete training for the *PCI* group). As noted, these teachers were either unable to attend the initial training or joined the study afterwards.

PCI Materials

The *PCI Reading Program - Level One* is a sight word based program designed to help non-readers become successful readers. The curriculum was developed specifically for students with developmental disabilities, autism, and significant learning disabilities. Because it is a mastery-based, individualized program, students can learn at their own pace. The program is also multi-sensory based, so students can use various cues and manipulatives to help them learn. The program aims to teach students 140 sight words and common nouns and verbs through visual discrimination. The recommended implementation of the program specifies a system of repetition, practice, errorless discrimination, controlled reading, and high-interest activities. Specifically, students learn through a series of steps including learning the word, tracing the word, hands-on practice, independent practice, and repetition of these steps. Next come review, assessment, and, finally, reading a book.

The complete program contains word building lessons, supplemental lessons and activities, guided word practice, a trace-and-read workbook, flashcards, and a word viewer. Also embedded in the program are periodic assessments for teachers to administer as part of the learning cycle. Teachers are supplied with a teacher's guide and a checklist for student progress. The program includes reproducible sheets for parents to work on with their children.

District Materials

Background information obtained from the teacher consent forms and surveys showed that, prior to this study, teachers across both districts used a variety of materials for reading instruction. Neither district truly had a central district curriculum. The most popular programs were EdMark, Houghton Mifflin, The Letter People, and READ 180, although many more instructional sources were noted. In many cases, individual teachers reported using more than one source for reading instruction and several teachers actually reported "none" as their main curriculum.

Expectations of Implementation

Expectations for implementation were discussed and agreed upon during the individual district trainings. *PCI* teachers are expected to use *PCI* as their core reading program for all participating students. Teachers are expected to follow the curriculum and lesson cycles directly, as outlined in the Teacher's Guide. At a minimum, each student should receive 20 minutes of *PCI* instruction per day, in order to comply with the publisher's definition of minimum acceptable implementation. According to *PCI*, ideal implementation is considered to be about 45 minutes per day, the rate at which a "typical" participating student will complete the program within one school year. *PCI* group teachers at both sites verbally agreed to teach *PCI* for their full reading block, which lasts about 45 minutes per day.

Schedule of Major Milestones

Table 1 lists the major project milestones and associated dates.

Table 1. Research Milestones

Milestone	Date
April 2007	Initiation of the experiment
June to July 2007	Recruitment of school districts and assessment developer
August 29, 2007	Development of assessment and district approval
September to October 2007	Question and answer sessions, randomization and training
October 2007	Administration of pre-assessments, start of implementation, and initiation of monthly web surveys
April 2008	Classroom observations
May 2008	Administration of post-intervention assessments and completion of data collection

Participant Recruitment and Site Descriptions

How the Sample was Identified

The way participants for a study are chosen largely determines how widely the results can be generalized. In this case, the sample was identified based on the number of teachers who met the criteria of teaching students who 1) have severe disabilities, 2) were in 3rd through 8th grade, and 3)

were taught in a self-contained reading block. The *PCI Reading Program - Level One* is tailored to non-readers with developmental disabilities, autism, and significant learning disabilities. Therefore, we were careful to include in the study *only* students with severe disabilities. Initially, we attempted to limit our study to non-readers in grades 3-5. However, in order to increase the sample size, we ultimately decided to extend our sample to include middle school grades as well. Self-contained reading blocks were set as a requirement for participation in order to ensure a measurable framework for implementation time that would be comparable across grades and various classroom settings.

The existing relationship between *PCI* and the administrative staff at the Brevard Public Schools Special Education Department aided in our recruitment of teachers to participate in the research within the district. Together with the person designated as *PCI*'s Point of Contact (POC), we contacted the district resource teacher to explain the details and procedures for the study. She agreed to act as the district POC for the study and identified eligible teachers who met the criteria.

Researchers hosted a voluntary telephone question-and-answer session for all interested teachers. The majority of identified teachers called in to the meeting, which provided a format for us to describe the specifics of participation in the study as well as to answer their questions and to address their concerns. All identified teachers were then invited to an all-day training. Thirteen teachers attended this initial training for the research experiment on September 25, 2007.

Our BPS Point of Contact referred us to the Miami-Dade County Public Schools Division of Special Education. The BPS POC works closely with the MDCPS Instructional Supervisor for Programs for Students with Mental Handicaps and Physical Impairments, who agreed to act as POC for the study in MDCPS.

The MDCPS POC sought to identify teachers who met the criteria for the study. She conducted an extensive process of teacher recruitment by contacting program specialists and site directors. In addition, several teachers were recruited at training sessions for the Florida Alternative Assessment. Aspiring participants were invited, through the district POC, to attend one of two question-and-answer telephone sessions with us. Only a few teachers attended either meeting.

The research design required a recruitment of 35 to 40 teachers in MDCPS in order to fulfill a total sample size of 50 teachers across both districts. Due to the timing of teacher recruitment, the beginning of the study was delayed. Consequently, randomization was conducted on three separate occasions. The initial teacher training for the research study in MDCPS occurred on October 18, 2007 with 27 teachers.

Brevard Public Schools

Brevard Public Schools (BPS) serves Brevard County, Florida, and is based in the city of Viera. Brevard County is a large suburb located approximately 50 miles southeast of Orlando. The total population of the county was estimated to be 534,359 in 2006 (U.S. Census Bureau, 2000).

BPS has 126 schools serving pre-kindergarten through grade 12. The total enrollment is 74,791 students (Florida Department of Education, 2006). Table 2 provides information about the entire district.

Table 2. Demographics of Brevard Public Schools

Brevard Public Schools	
Total schools	123
Total teachers	5,120
Grade structure	PK-12
Student enrollment	73,369
Percent of students designated as:	
Disabled	16.8%
English language learner	2.4%
White	69.7%
Black	14.7%
Hispanic	8.1%
Asian/Pacific Islander	2.0%
American Indian/Native Alaskan	0.3%
Multi racial	5.2%

Source: Florida Department of Education, 2007

Miami-Dade County Public Schools

Miami-Dade County Public Schools (MDCPS) encompasses Miami, Florida, and the city's surrounding suburbs. The county's total population was estimated to be 2,402,208 in 2006 (U.S. Census Bureau, 2000).

MDCPS has 441 schools serving pre-kindergarten through grade 12. The district's total enrollment is 353,783 students (Florida Department of Education, 2006). Table 3 provides information about the entire district.

Table 3. Demographics of Miami-Dade County Public Schools

Miami-Dade County Public Schools	
Total schools	471
Total teachers	23,629
Grade structure	PK-12
Student enrollment	348,128
Percent of students designated as:	
Disabled	11.4%
English language learner	15.5%
White	9.2%
Black	26.3%
Hispanic	62.0.%
Asian/Pacific Islander	1.2%
American Indian/Native Alaskan	0.0%
Multi racial	1.3%

Source: Florida Department of Education, 2007

Randomization Schedule in the Two Districts

Randomization of our teachers at the Brevard Public Schools occurred on September 18, 2007. Thirteen teachers submitted Teacher Agreement Forms, consenting to participate in the study. We matched teacher pairs on grade level, use of reading program, teaching experience, and number of students in each class. To ensure a balanced distribution, six pairs of teachers were assigned via coin toss either to the *PCI* or to the control condition. The thirteenth teacher was assigned using a coin to the *PCI* condition. Although we had received additional information on three of the teachers beyond of the scope of our standard questions, these teachers were matched independently of this information.

The initial randomization of our teachers at the Miami-Dade County Public Schools occurred on October 11, 2007. This deadline was established in order to allow the district POC sufficient time before the training to notify teachers about the group assignment meeting and to secure substitute teachers for all participants. Twenty-six teachers had submitted Teacher Consent Forms by this date. Matched pairs were based on whether teachers taught a self-contained class, grade level, and total teaching experience. Thirteen pairs of teachers were assigned by coin toss either to the *PCI* or to the control condition, to ensure a balanced distribution.

After the initial deadline, four additional teachers submitted consent forms. In order to accommodate these teachers and increase our sample size, we randomized a second cohort of teachers on October 15, 2007. The four teachers were paired on grade level, school, and program used. They were added and assigned prior to the scheduled *PCI* training on October 18, 2007.

During the MDCPS training, the district POC informed us of two additional teachers interested in the study. We established a final deadline for participation and this third teacher cohort was randomized on October 23, 2007. Teachers in each of 22 pairs were randomly assigned to the *PCI* or control condition using a coin toss. Three additional teachers, where appropriate matches could not be made, were randomized to condition individually through a coin toss.

Data Sources and Collection

The data for this experiment are primarily those provided by the school districts and collected by Empirical Education. They consist of student pre- and post-intervention sight word assessment scores, phonological assessment data, student demographic data, and data from teacher emails, surveys, training, and classroom observations. In addition, we have reviewed various program documents and materials. Therefore the research employs a multiple methods approach through which we measure and document the intervention implementation to provide qualitative and quantitative descriptions of the program.

District-Supplied Information

Researchers requested records and other background information for the students who were taught by participating teachers. Specifically, the districts were asked to provide the following data:

- Student name or unique ID
- Gender
- National School Lunch Program status (proxy for socio-economic level)
- Ethnicity
- English learner status
- Date of birth
- Grade
- Classroom teacher
- School the student attends

All student and teacher data having any individually identifying characteristics were stripped of such identifiers, and the data were stored using security procedures consistent with the provisions of the Family Educational Rights and Privacy Act (FERPA).

Empirical Education received class rosters from BPS in October 2007 and from MDCPS in December 2007.

Achievement Measures

Sight Word Pre- and Post-Intervention Assessment

The primary outcome measures are student assessment scores on the Sight Word Post-Assessment, which was developed by an independent consultant. The development specialist took the following steps to determine the appropriate words for both the pre- and the posttests.

1. Selected only words that are taught in the *PCI* program, with 50% of those words being common to the EdMark reading program.
2. Used the EDL Reading Core Vocabulary Cumulative list to determine the reading levels of each word. This was important so that both the pre- and post-assessment had an even distribution of words at the primer level and the first-grade level.
3. Used the Brown Corpus frequency list to determine the frequency of each word. It was important to have an equal distribution of more and less frequently used words. Brown Corpus determines the frequencies in percentages and instances. For example, the word "the" has nearly 7,000 instances with a frequency of 6.89%.
4. Divided the resulting word list into quartiles based on when the words are introduced in each of the two programs. Introduction is as important as frequency when determining the words for the tests. For example, the word "it" was presented as word 69 in the *PCI* reading program and word 64 in the Edmark program. Thus "it" was an appropriate word to select for the test because of the similar introduction in both programs. By having a distribution of words introduced in the beginning, middle, and end of the program, any memory issues are ruled out.

These steps were taken in order to rule out any variances of primer and first-grade words, infrequent versus frequent words, and any long term memory issues. The result was a sight word pre-assessment and post-assessment of 20 words each, with only one common word across the two. Each assessment, while technically not timed, was to take approximately 6-8 minutes per student. Administrators of the test were instructed to document any modifications to display and wait time based on individual student needs. Additionally, test administrators were required to record whether students answered the question correctly the first time, were able to self-correct their answers, chose to pass, or answered incorrectly. A test of Kuder-Richardson Formula 20 (KR20) showed a very high reliability for the test items.

Phonological Assessment

In addition to the Sight Word Assessments, a Phonological Assessment was also developed by an independent consultant. The Phonological Assessment, administered to all participating students at the beginning of the study, includes three sections: Recognition of Sounds, Initial Consonant and Vowel Sounds, and Ending Consonant and Vowel Sounds. Each section consists of five questions, for a total of 15 questions across the assessment. In scoring the Phonological Assessment, teachers were required to qualify student answers by reporting whether students answered the question correctly the first time, were able to self-correct their answers, required assistance from the teacher, chose to pass, or answered incorrectly. The Phonological Assessment is not used as an outcome measure, since *PCI* does not aim to teach phonics. Instead, the Phonological Assessment is used to provide additional background information on participating students' reading skills and is examined as a moderator to assess whether outcomes differ for students who begin the program at various levels of pre-reading skills.

Pre-Assessment Training

Although many participants had previous experience in administering sight word assessments, teachers were not familiar with the specific administration of the assessments described above. In the pre-assessment training, expectations of the program and the study were reiterated both for pilot and for control teachers. The majority of the assessment training was led by a researcher from Empirical Education. It was made clear to *PCI* teachers that the pre-assessments must be administered before instruction in the *PCI* program begins.

The BPS training occurred on September 25, 2007 with all 13 participating *PCI* and control group teachers in attendance. The district POC also attended and helped to answer teacher questions specific to the district context. Although the teachers had been randomized prior to the training, they had not yet been notified of their group assignments. Therefore, the training began with the researcher announcing teacher assignments. Those assigned to the *PCI* group appeared to be pleased. In BPS, the day-long training started with the pre-assessment training. Upon completion of this session, control teachers were dismissed while the *PCI* teachers remained for the full training.

MCDPS teachers had been notified of the randomization assignments by the district POC prior to the training. A few teachers were unable to attend the initial *PCI* training due to a pre-existing district conflict. A total of 27 teachers attended at least one portion of the training. Several teachers in the control group had a pre-existing commitment that would not allow them to participate in any training until the close of the school day. In order to accommodate and include a larger number of teachers in the training, the pre-assessment training was administered in the afternoon, after the completion of the *PCI Reading Program - Level One* training.

In both districts, the pre-assessments were given between October and December 2007, and all post testing was conducted in May 2008. Teachers were instructed to return the completed assessments in postage-paid envelopes

Methods Used to Investigate the Intervention Implementation

In addition to pre-test, post-test, and demographic data, we also collected implementation data over the entire period of the experiment, beginning with the randomizations and ending with the academic calendar of the districts in May 2008. Training observations, classroom observations, informal interviews, multiple teacher surveys, email exchanges, and telephone conversations are used to provide both descriptive and quantitative evidence of the implementation.

Observational Data

In general, observational data are used to inform the description of the learning environment, instructional strategies employed by the teachers, and student engagement. These data are minimally coded. We observed the BPS training on September 25, 2007 and the MDCPS training on October 18.

Classroom observations took place in both districts during late April of 2008. While the scheduling of classroom observations was tailored to convenience, the sample of classrooms selected for observation represented the various contexts existing within this study. Our goal was not only to visit the classrooms of nearly one-third of participating teachers, but also to meet with a small sample of school principals. Class selection for observations was based on "stratified convenience." Classes were selected for observations to represent an equal number of middle and elementary schools as well as an equal number of *PCI* and control classes.

In MDCPS, we first selected schools containing multiple teacher participants in order to maximize the number of teachers observed within the time allotted. Next, we looked at whether we had a fair representation of both middle and elementary schools in our sample. Following, we made sure to observe equal numbers of teachers in both the control and *PCI* groups. The remainder of schools observed, those without multiple teacher participants, was selected based on convenient location and to achieve balance for either grade level of school or assignment to condition.

In BPS, there were only two schools with multiple teacher participants, both of which had all *PCI* teachers. Only one of these schools was selected for observations based on the convenience of its location. The remaining schools for this district were selected based on grade level of school or condition.

Once the sample of classes was identified, participating teachers in each selected school were contacted to obtain information about their class schedules. Observation times were scheduled based on the time teachers taught reading.

At Miami-Dade, we visited 4 out of 19 schools and 9 out of 28 classes. At Brevard, we visited 4 out of 11 schools and 5 out of 12 classes. This selection allowed us to visit 8 out of 20 schools and 14 out of the 40 classes in the study.

During these classroom visits we observed how teachers designed and carried out instruction. We were also interested in how teachers organized instruction—group work, individual work, one-on-one instruction—across both assignment groups. Specifically, in *PCI* classrooms, we documented teachers' use of materials as well as how closely they followed the prescribed curriculum. For the control group, we hoped to obtain a glimpse of the different curricula enacted across the classrooms and to understand the degree to which instruction is individualized for students. All classroom observations were conducted within a period of one week across both districts.

Survey Data

The quantitative survey data are reported using descriptive statistics; these are summarized by individual teacher and by assignment group (*PCI* and control), and are compared by group. The free-response portions of the surveys are minimally coded. Survey data are used to quantify the extent of exposure to the materials, that is, opportunities to learn with the curriculum.

Surveys were deployed to both *PCI* and control group teachers beginning in October 2007 and continued on a monthly basis through May 2008.

Surveys covered several topics:

- Teacher Background and Classroom Context
- Training
- Assessment
- Student Progress
- Tasks, Activities, and Materials
- Instructional Time
- Teacher Satisfaction
- Student Engagement

Table 4 shows the response rates for the 21 control teachers and 23 *PCI* teachers participating in the study. The response rates were extremely high, with an overall rate of 93% for all surveys combined.

Table 4. Survey Response Rates

Survey Number	Date	Control Response Rate	*PCI* Response Rate	Overall Response Rate
Survey 1	October 26, 2007	90.5%	100.0%	95.5%
Survey 2	November 30, 2007	95.2%	95.7%	95.5%
Survey 3	January 11, 2008	100.0%	91.3%	95.5%
Survey 4	February 8, 2008	95.2%	91.3%	93.2%
Survey 5	March 14, 2008	95.2%	95.7%	95.5%
Survey 6	March 28, 2008	95.2%	82.6%	88.6%
Survey 7	April 25, 2008	90.5%	91.3%	90.9%
Survey 8	May 16, 2008	90.5%	87.0%	88.6%

The survey topics were developed to account for the various aspects of teacher actions associated with instruction and learning. In order to characterize the extent of implementation, we used a repeated question strategy to document the average time teachers spend on specific activities. Because strict adherence to the Teacher's Guide is specified by *PCI* Education and supported by the researchers, we tracked how closely teachers adhered to the lesson cycle in their classrooms. In certain cases, we asked identical questions across multiple surveys in order to develop an adequate average as well as to gain an understanding of variation at different times during the school year.

Teacher Background and Classroom Context

The first survey asked questions about the teacher's background and classroom context. Teacher background questions included education (level completed and major area of study), years of teaching experience (overall and Special Education), and credentials and certification. Because recent literature correlates teaching experience and content knowledge with teacher quality, we examined whether the impact of *PCI* was greater or less for teachers with more experience teaching Special Education (Amrein-Beardsley, 2006; The Center for Public Education, 2005).

We also asked very basic questions about classroom organization. For example, we asked how students are organized while instruction is delivered; e.g., one-on-one instruction, group instruction, or independent student work. For *PCI* specifically, we also asked how students are typically organized for each part of the lesson cycle. Because the *PCI* curriculum specifies which aspects of the program are to be delivered in direct one-on-one instruction, which parts can be taught to a group, and which exercises may be completed independently by students, the answers to these questions can help inform how closely teachers adhered to the Teacher's Guide. Moreover, we asked questions about who is delivering each piece of instruction. Many of our participating classes receive assistance from paraprofessionals, aides, parents, or other adults. The publisher maintains that the *PCI* program can be taught by any adult familiar with the program, and we are interested in learning *who* is teaching the program in participating classrooms.

Training

PCI was particularly interested in learning how teachers felt about the effectiveness of the *PCI* training and whether each component of the training was necessary to enable teachers to implement the program well. Moreover, although the *PCI* program does not include any follow-up training, it is important to understand the extent to which *PCI* and control teachers received additional support and training from their districts throughout the school year.

Assessment

Assessment is an important component of *PCI*'s mastery-based curriculum. It is important for us, as researchers, to know whether teachers were assessing students in the manner the program instructs and whether they were using assessment results to differentiate instruction. This again helps inform the extent of implementation. Teachers were also asked if they used other assessments outside of the *PCI* curriculum. Before the state testing period began, teachers were asked if instruction changed during that period and if they supplemented the *PCI* program in any way in order to prepare for standardized testing.

Student Progress

In addition to questions regarding assessment, teachers were asked periodically about the progress of their students. Because the *PCI* program is individualized and not all students could be expected to move through it at the same pace, we asked teachers where the majority of students were in the curriculum and about the progress of students who moved at the fastest and slowest rates. The goal of the *PCI Reading Program - Level One* is for students to master more than 100 words. Because the pre- and post-tests are sight word assessments, the amount of content covered will likely affect performance on the post-assessment. Therefore, we wanted to know how many words students had learned at different points in the school year. Additionally, we can examine how students progress through the program in relation to the amount of time the teacher spends on *PCI* instruction.

Tasks, Activities, and Materials

When we asked questions about the lesson cycle, we not only asked who teaches each step and how instruction is organized, but we also asked about the amount of time spent on each step and how regularly each step is completed. Again, *PCI* teachers are expected to adhere strictly to the lesson cycle. This is stated in the curriculum and was reinforced during training. From teachers' answers to questions about the lesson cycle, we also gained an understanding of how much time a "typical" student spent on *PCI* reading each week. Similar questions about time on task were asked of teachers in the control group.

In addition to asking about use of the mandated *PCI* materials, we also asked questions about bonus materials. The Activity Sheets are a required component of the program, but these worksheets are reproducible and may be sent out as homework, which is not required. For researchers, asking questions about homework also helps inform our measures of time spent on task. For the publisher, reactions to supplemental materials such as the CD-ROM and the Building Reading Skills binder provide valuable feedback on what the teachers find useful about the program. Each of the first few surveys asked whether the teachers had all the materials needed to fully implement the program. This information, in addition to pre-assessment and other survey data, tells us how many months of the year teachers actually spent implementing *PCI*.

Instructional Time

Each month participating teachers were asked to record how many minutes their "typical" student received *PCI* instruction during a period of one week. Teachers were asked to break down this instruction into subgroups: how much of this time did the teacher lead instruction and during how much of this time was instruction led by another adult (aide, paraprofessional, parent, other adult), how much time was spent on one-on-one instruction and how much of this time was instruction completed in a group setting or independently. Survey responses not

received within three weeks of the time period specified in the survey were not included in the analysis.

Teacher Satisfaction and Student Engagement

The surveys also asked questions about student engagement and teacher satisfaction with the *PCI Reading Program*. During the training, several MDCPS teachers complained of not having *anything* engaging to do with their students prior to *PCI*. Because student engagement is an aspect of the program that is critical to the teachers, it is an important element to measure. During the first months of implementation, we received many informal reports from teachers regarding the extent to which students were enjoying the program. Therefore we wanted to measure this engagement through surveys, especially in comparison to students in control classes. Additional questions probed challenges teachers may have experienced, preferences and possible modifications, and plans for future implementation.

Formation of the Experimental Groups

This section describes the sample that we will use to determine the impact of *PCI*. The sample consists of teachers and students, where teachers have been randomly assigned to *PCI* or the control condition. We describe this sample as being formed initially through the random assignment but modified somewhat through attrition or loss of units at different points during the experiment for a variety of reasons. Ideally, by randomizing assignment into the two conditions, the groups should look the same in terms of important characteristics such as demographic composition, prior achievement, and teacher characteristics. In addition because we paired teachers, we can expect somewhat better balance than we would have if we hadn't first balanced them on these characteristics. However, by chance (as well as the imprecision of the pairing) the groups are never exactly balanced and may differ on important characteristics likely to affect the outcome. Furthermore, the loss of teachers and students during the experiment may cause the observed difference between conditions to reflect imbalance in the sample instead of differences caused by exposure to *PCI* versus the control. This would happen, for example, if teachers are more likely to drop out of the program than the control group because of the extra burden.

Therefore in this section we inspect the distribution of teachers, classes, grades, and students, looking in particular at the balance between the *PCI* and control groups. We look at whether there was differential attrition between the *PCI* and control groups both overall and with respect to subgroups of students and teachers. We also inspect the final sample that is available for determining impact and check whether the *PCI* and control groups are balanced on important characteristics, recognizing that imbalance may have entered into the sample both because of "unlucky" randomization and through differential attrition. (For this accounting, we focus on the data available for Sight Word assessment results which we consider the primary outcome measure.)

Number of Units in the Sample and Attrition

Table 5 shows changes in the sample from the point at which the teachers were randomized to the point at which the posttests were received.

Table 5. Numbers of Units in the Experimental Groups and Attrition Over Time

Event	Control		*PCI*	
	No. of teachers	No. of students	No. of teachers	No. of students
Randomization	23	n/a	24	n/a
(Loss prior to rosters)	(5)	n/a	(2)	n/a
Fall rosters received	18	105	22	87
(Loss before/at pretest)	(2)	(29)	(2)	(10)
Pretest scores received	16	76	20	77
(Loss before/at posttest)	(1)	(15)	(0)	(10)
Final count of units with pretest and posttest	**15**	**61**	**20**	**67**

Explanation of Teacher Attrition

By January 2008, a total of three teachers had left the study after being randomized. One control teacher from MCDPS was unable to attend the initial training due to a family emergency and subsequently decided she/he would be unable to participate in the study. A second MCDPS teacher, who was randomized after the initial *PCI* training and assigned to the *PCI* group, withdrew from the study because she/he discovered during the make-up training that her/his students were not at an appropriate reading level for the study. In December, after receiving the student rosters from BPS, we excluded a control teacher from the remainder of the study because we discovered that she/he had no students that met the requirements for participation. As described in the explanation of student attrition, five additional teachers were unable to provide the required parental consent for any of their students and these also could not be used in the study. Final participant totals are reflected in the tables above.

We see from Table 5 that at the start of the experiment, 24 teachers were assigned to *PCI* and 23 teachers were assigned to the control condition. After limiting the sample of teachers to only those who provided fall roster information, and for whom we received pretest and posttest scores for at least some students, the final number of teachers used in the analysis are 20 *PCI* and 15 control teachers. The number of students for this sample of teachers was 67 and 61 for *PCI* and control, respectively.

Explanation of Student Attrition

We ran into several obstacles in collecting student roster data in MDCPS. The district requires parental consent for the release of student data. District consent forms were provided to us and handed out to teachers at the training. The district form was also made available in Spanish and Creole translations. Upon receiving signed parental consent forms, teachers were asked to fax the forms to a district contact. The contact at the district then provided student data to us for all students for whom they had signed consent forms.

By the beginning of January, we had received parental consent forms from 27 (13 pilot, 14 control) of the 32 MDCPS study teachers. Many attempts were made to contact teachers by both researchers and our district contacts. A few teachers proved to be non-responsive to any communication. Several teachers expressed concern over the fact that the forms had been sent home several times and were never returned. In BPS, no parental consent form was required, as the district considers research to be within the realm of everyday instruction. Therefore, we were

able to include in the analysis data for any students for whom we received assessments. We received student roster data for all 12 participating teachers. We provide total counts in the cases for which rosters were received in Table 5.

Students are listed as attrition if they were missing a pretest score (5 in the *PCI* group and 27 controls) or if the roster information was not received from the district.

In addition, we excluded students for whom we may have received a pretest but were considered ineligible because they were (1) designated as being non-verbal (4 in *PCI* and 14 in the control condition), or (2) obtained a perfect score of 20 on the pretest (14 in *PCI* and 9 in the control condition).

If our goal is to estimate the impact of exposure to *PCI* compared to the control experience on student performance, then we would like to maintain the equivalence between groups that we obtain originally through randomization. Otherwise the differences that we measure at the end of the experiment don't just reflect differences in classroom experiences of students, but they reflect other effects that lead to different compositions of cases in the two conditions. For this reason we examine whether there is reason to think that special categories of students were systematically excluded from the two conditions at different rates, and whether students belonging to those categories are expected to perform above or below average. We observe the following patterns:

1) More controls than *PCI* students are missing pretests; however, it is not clear that student who are missing pretests would perform systematically higher or lower than those with pretests (if students with missing pretests were not tested because the assumption is that they would perform especially poorly, we would expect a difference on pretest performance between the *PCI* and control condition for students who do have a pretest (since a larger proportion of these presumed low performers would have been removed from the control condition); however, we don't see this result, as detailed below.)
2) We don't expect students who score 20 to be selectively excluded at different rates from the two conditions. Students in both conditions who score 20 are automatically excluded. Differences between conditions in proportions of such students are likely due to chance.
3) We don't expect students who are deemed non-verbal to be selectively excluded at different rates from the two conditions. Presumably the designation is established independently of students' assignment to conditions and students in either condition who are deemed non-verbal are automatically excluded. Differences between conditions in proportions of such students are likely due to chance.
4) We do not believe that the probability of receiving roster information is associated with whether a student is assigned to *PCI* or the control; therefore, loss of students due to their exclusion from the roster is likely due to chance only.

Randomization assures us that the groups formed are statistically equivalent – they differ only by chance – which allows a fair test of the impact of an intervention. When cases drop out for reasons other than chance then the impact that we measure may reflect differences in composition between the two groups that have a bearing on performance rather than differences in the programs received. Although we exclude from analysis student outcomes for teachers who either dropped out or did not supply roster or test information, the tests of balance in Table 10, and the test of differential attrition which is discussed below, show that this exclusion of cases did not lead to a large imbalance between conditions in the student sample that would result in differences of performance. This means that our subsequent estimates of the effects of *PCI* are unlikely to reflect differences in the composition of the *PCI* and control groups.

Differential Attrition

We also examine the rate of differential attrition for the 153 students included in the balance checks reported in Table 10. If there is a difference between *PCI* and control in the proportion of students with posttests, and if there is a difference between students with and without posttests in how they perform on the pretest then we expect that the differences between *PCI* and control on the posttest

reflects the differential loss of posttests. In Table 6 we see that the difference between conditions in the proportion of students with posttests is small and is easily due to chance. Also, in Table 7 we see that there is little difference in how students with and without posttests perform on the pretest. Either of these outcomes is sufficient to show that our impact estimates are not affected by differential attrition.

Table 6. Missing Tests for *PCI* and Control Groups

	Missing Tests		
Condition	**Pre & posttest**	**Pretest only**	**Totals**
Control	61	15	**76**
PCI	67	10	**77**
Totals	**128**	**25**	**153**
Statistics		**Value**	***p* value**
Fisher's exact		0.09	.28

Table 7. Difference in Pretest Scores for Students Having Pre- and Posttest Scores Versus Pretest Only

Descriptive statistics: Pretest outcomes	**Raw group means**	**Standard deviation**	**Number of students**	**Standard error**	**Effect size[a]**
Have pretest scores only	4.12	5.67	25	1.13	-0.18
Have both pre- and posttest scores	5.17	5.80	128	0.51	
***t* test for difference between independent means**	**Difference**		**DF**	***t* value**	***p* value**
(Missing posttest) – (Have posttest)	-1.05		151	0.83	.41

[a] The difference we are measuring is not an effect of treatment (the usual sense of effect size) but a result of chance differences in the randomization.

Differences in Teacher Perception of Pretest Difficulty

In one of the surveys, we asked teachers about the difficulty of the pretest assessments in relation to student ability. We offer these results here as additional information as to a possible discrepancy between the *PCI* and control groups. We expected the assessments to be difficult for this group of students. All respondents who marked "other" described the test as being hard for some students and easier for others. Regardless, these data show that the control teachers overwhelmingly considered the test more difficult for their participating students than did the *PCI* teachers. We observed similar responses with regard to the Phonological Assessment.

Table 8. Was the Sight Word Pre-Assessment At the Appropriate Difficulty Level For Your Participating Students?

	Yes	No, too difficult	No, too easy	Other
Control (N=17)	11.8%	76.5%	5.9%	5.9%
***PCI* (N=21)**	71.0%	14.3%	4.8%	9.5%

Note. Fisher's exact test gives p value <.01

Table 9. Was the Phonological Assessment At the Appropriate Difficulty Level For Your Participating Students?

	Yes	No, too difficult	No, too easy	Other
Control (N=17)	29.4%	64.7%	5.9%	0.0%
***PCI* (N=21)**	66.7%	23.8%	0.0%	9.5%

Note. Fisher's exact test gives p value =.01

As we see in the next section, the two groups did not differ significantly in their pretest scores. It may be the case, however, that control teachers had more students who they considered ineligible and therefore did not test.

Characteristics of the Initial Sample

In Table 10 we compare the composition of the control and *PCI* groups at the point we received the rosters[2]. For each of the characteristics of this sample, we conducted a statistical test [3] to determine the likelihood of obtaining a chance imbalance as large as or larger than the one observed. Of course the randomization assures us that any initial imbalance is a result of chance, and is not an indication of selection bias; however, some attrition did occur between the time of randomization and the point at which we received the rosters, therefore it is useful to examine the groups that we received information on to see whether the amount of imbalance is something we would expect to see by chance. We see that the *PCI* group had a higher portion low social economic status students and the group had a higher average pretest score; however, the high *p values* indicate that such differences can easily result from chance.

Table 10. Characteristics of Study Sample

	Control group	*PCI* group	Less than 5% chance of seeing this much imbalance
Student characteristics			
English proficient	75 (98.68%)	73 (94.80%)	No (p =.27)
Low socio-economic status	56 (73.68%)	60 (77.92%)	No (p =.82)
Mean sight word pre-test score	4.26	5.73	No (p =.32)
Teacher characteristics			
Less than 4 years special education teaching experience	2 (12.50%)	4 (20%)	No (p = .67)

[2] Our balance checks involve 153 students who are included on the roster and for whom we have pretest scores. These students belong to the classes of the 20 *PCI* teachers 16 of the 15 control classes that get included in the final analysis (one additional control class is excluded because we did not receive any posttest scores for that teacher.) Therefore, our balance checks apply to a sample of students and teachers who are very close to the sample used in the final analysis. Importantly, we will examine whether the pretests scores for students in the two conditions are balanced following attrition – balance on this factor gives us assurance that the final impact does not simply reflect preexisting differences.

[3] For the categorical variables, we used the Fisher Exact Test and for the continuous variables we used a *t* test adjusted for clustering. In both cases the criterion for significance was set at <.05.

Distribution by Grade

Table 11 shows the distribution by grade of the 153 eligible students for whom rosters and pretests were received.

Table 11. Distribution of Students By Grade

	Grade Level							
	2	**3**	**4**	**5**	**6**	**7**	**8**	**Total**
Control	3	10	8	1	17	19	18	**77**
PCI	0	10	12	12	11	16	16	**76**
Total	**3**	**20**	**20**	**13**	**28**	**35**	**34**	**153**

Statistical Equations and Reporting on the Impact of *PCI*

Setting Up the Statistical Equation[4]

We put our data for students and teachers into a system of statistical equations that allow us to obtain estimates of the direction and strength of relationships among factors of interest. The primary relationship of interest is the causal effect of the program on a measure of achievement. We use SAS PROC MIXED (from SAS Institute Inc.) as the primary software tool for these computations. The output of this process are estimates of effects as well as a measure of the level of confidence we can have that the estimate is true of the hypothetical population to which the experiment is meant to generalize.

Program Impact

A basic question for the experiment was whether, following the intervention, students in *PCI* classrooms had higher reading scores than those in control classrooms. Answering this is not as simple as comparing the averages of the two groups. The randomization gave us two groups that are equivalent to each other on average in every way, except that one receives *PCI* and the other one does not. But as we saw in the section on the formation of the experimental groups, in a single randomization we expect chance imbalances. By using the pretest in the equation, we account for some of the extraneous variation in the outcome, which increases the precision of our estimate of the effect of *PCI*. We also have to account for the fact that students are

[4] The term 'statistical equation' refers to a probabilistic model where the outcome of interest is on the left hand side of the equation and terms for systematic and random effects are on the right hand side of the equation. The goal of estimation is to obtain estimates for the effects on the right hand side. Each estimate has a level of uncertainty which is expressed in terms of standard errors or *p* values. The estimate of main interest is for the treatment effect. In this experiment, we model treatment as a fixed effect. With randomized control trials, the modeling equation for which we are estimating effects, takes on a relatively simple form: Each observed outcome is expressed as a linear combination of a treatment indicator, one or more covariates that are used to increase the precision of intervention effect, and usually a series of fixed or random intercepts, which are increments in the outcome that are specific to units. As a result of randomization, the other covariates are distributed in the same way for both the treatment and control groups. For moderator analyses we expand these basic models by including a term that multiplies the treatment indicator with the moderator variable. The coefficient for this term is the moderator effect of interest.

clustered by classes and teachers. We expect outcomes for students who are in the same class or who have the same teacher to be dependent as a result of shared experiences. We have to add this dependency to our equation or else our confidence levels about the results will be artificially high.

Covariates and Moderators at the Student and Teacher Level

In addition to estimating the average impact, we also include in the equation other variables (called covariates) associated with characteristics of the students and teachers, which we expect to make a difference in the outcomes for the students. For example, as was described above, we add the pretest score into almost all our statistical equations in order to increase precision. In addition, we consider whether there is a difference in the effect of the intervention for different levels of the covariates. For example, we consider whether the program is more effective for higher-performing students than for lower-performing students. We estimate this *difference* (between subgroups) *in the difference* (between the program and control groups) by including an interaction term in the statistical equation. This term multiplies together the variable that indicates whether the student is in the intervention group, and the covariate. We call covariates, that are included in such analyses, potential "moderators" because they may moderate—either increase or decrease—the effect of the program on student outcomes. The value for the interaction term is a measure of the moderating effect of the covariate on the effect of the program.

Fixed and Random Effects

The covariates in our equations measure either 1) fixed characteristics that take on a finite set of values (e.g., there are only two levels of gender); or 2) a set of characteristics that is assumed to have a distribution over a population and where we treat the values that we measure as though they were a random sample from that larger population. The former are called "fixed effects", the latter, "random effects". Random effects add uncertainty to our estimates because they account for sampling variation, or the changes we would observe in the outcomes if we re-sampled units from the same hypothetical population. Fixed effects produce less uncertainty but also limit the extent to which we can generalize our results.

We usually treat the units that were randomized as "random effects", so that in the statistical equations, our estimates reflect the degree of uncertainty that comes if we were to draw a different sample of units from the same population[5]. This allows us to argue for the generalizability of our findings from a sampling perspective. Treating the units that were randomized as fixed, forces us to use other arguments if our goal is to generalize.

Using random or fixed effects for participating units serves a second function—it allows us to more accurately represent the dependencies among cases that are clustered together (e.g., students in classes.) All the cases that belong to a cluster share an increment in the outcome--either positive or negative—that expresses the dependencies among them. An appropriate measure of uncertainty in our estimate of the program's effectiveness takes into consideration the relative levels of variation *within* the larger units and *between* them. All of our statistical equations include a student-level error term. The variation in this term reflects the differences

[5] Although we seldom randomly sample cases from a broader population, and in some situations we use the entire population of cases that is available, we believe that it is still correct to estimate sampling variation (i.e., model random effects). It is entirely conceivable that some part or the whole set of participants at a level end up being replaced by another group (for whatever reason) and it's fair to ask how much change in outcomes we can expect from this substitution.

we see among students that are not accounted for by all the fixed effects and other random effects in our statistical equation.

The choice of terms for each statistical equation is not rigid but depends on the context and the importance of the factors for the question being addressed. The tables reporting the estimates resulting from the computation will provide an explanation of these choices in table notes where necessary for technical review.

Exploratory Investigations

Finally, to better understand unexpected results, we use other demographics, teacher characteristics, and supplementary observational data in exploratory investigations to generate additional hypotheses about which factors interact with the program. These results are considered exploratory because they often follow inspection of the results of analyses that are planned at the design stage of the experiment. Their primary goal is to inform future studies.

Reporting the Results

When we run the computations on the data, we produce several results: among them are effect sizes, the estimates for fixed effects, and *p* values. These are found in all the tables where we report the results.

Effect sizes

We translate the difference between program and control groups into a standardized effect size by dividing the average group difference by the amount of variability in the outcome. The amount of variability is also called the "standard deviation" and can be thought of as the average distance of all the individual scores from the average score (more precisely, it is the square root of the average of squared distances.) Dividing the difference by the standard deviation gives us a value in units of standard deviation rather than units of the scale used by the particular test. This standardized effect size allows us to compare the results we find with results from other studies that use different measurement scales. In studies involving student achievement, effect sizes as small as 0.1 (one tenth of a standard deviation) are sometimes found to be important educationally. When possible we also report the effect size of the difference after adjusting for pretest score and other fixed effects, since that adjustment provides a more precise estimate of the effect by compensating for chance differences in the average pretest of the program and control groups. Theoretically, with many replications of the experiment, these chance differences would wash out so we would expect the adjusted effect size on average to be closer to the true value.

Estimates

We provide estimates to approximate the actual effect size. Any experiment is limited to the small sample of students, teachers, and schools that represent a larger population in a real world (or hypothetical) setting. Essentially we are estimating the population value. When we report an estimate in a table, the value refers to the change in outcome for a one-unit increase in the associated variable. For example, since we code participation in the control group as 0, and participation in the program group as 1, the estimate is essentially the average gain that we expect in going from the control to the program group (while holding other variables constant).

p values

The *p* value is very important because it gives us a gauge of how confident we can be that the result we are seeing is not due simply to chance. Specifically, it tells us what the probability is that we would get a result with a value as large as—or larger than—the absolute value of the one observed when in fact there is no effect. Roughly speaking, it tells us the risk of concluding that the intervention has had an effect when in fact it hasn't. This mistake is also known as a "false-positive" conclusion. Thus a *p* value of .10 gives us a 10% probability of drawing a false-positive conclusion. This is not to be confused with a common misconception about *p* values: that they tell us the probability of our result being true.

We can also think of the *p* value as the level of confidence, or the level of belief we have that the outcome we observe is not simply due to chance. While ultimately depending on the risk tolerance of the user of the research, we suggest the following guidelines for interpreting *p* values:

1. We have a high level of confidence when $p \leq .05$. (This is the level of confidence conventionally referred to as "statistical significance.")
2. We have some confidence when $.05 < p \leq .15$.
3. We have limited confidence when $.15 < p \leq .20$.
4. We have no confidence when $p > .20$.

In reporting results with *p* values higher than conventional statistical significance, our goal is to inform the local decision-makers with useful information and provide other researchers with data points that can be synthesized into more general evidence.

Results

Teacher-Level Implementation Results

In this section we describe more fully the aspects of the implementation that characterize this intervention. Data for this section were obtained through surveys, classroom observations, and informal teacher interviews.

Conditions for Implementation

First we will present the data regarding the conditions that impact implementation in the classroom. We will describe teacher educational and teaching backgrounds, training, support, and the availability of materials and of other adults to help with instruction.

Teacher Background

At the beginning of the 2007-08 academic year, teachers were asked questions about their educational background and teaching experience. Provided in Table 12 are the responses regarding years of teaching experience in each program group as well as the number of years of experience each group has in teaching Special Education. The *PCI* group had more new teachers (17%) than did the control group (10%). There was an even bigger disparity in the number of teachers who were new to Special Education (see Table 13), with 26% among the *PCI* group compared to 10% among the control group.

Table 12. Years of Teaching Experience

	0-3	4-6	7-15	16+
Control (N=20)	10.0%	20.0%	15.0%	55.0%
***PCI* (N=23)**	17.4%	8.7%	39.1%	34.8%

Table 13. How Many Years Total Have You Taught Special Education?

	0-3	4-6	7-15	16+
Control (N=20)	10.0%	20.0%	20.0%	50.0%
***PCI* (N=23)**	26.1%	8.7%	34.8%	30.4%

All but two teachers in each assignment group had either a regular/standard teaching certificate or a specific certificate/specialization. Two teachers in each group had temporary certificates at the time of the survey.

Table 14. Teacher Credentialing and Certification

	Regular/ Standard	Temporary	NBPTS	Specific certificates for teaching bilingual, multicultural, limited English, or special education students	None	Other
Control (N=20)	75.0%	10.0%	0.0%	23.8%	0.0%	25.0%
***PCI* (N=23)**	91.3%	8.7%	8.7%	30.4%	0.0%	4.4%

Note. Teachers could select more than one category, so totals may exceed 100%.

A larger percentage of teachers in the control group obtained degrees in Elementary Education (20%) and Special Education (75%) than did teachers in the *PCI* group (17% and 52% respectively).

Table 15. What Was the Major Field of Study For Your Bachelor's Degree?

	Elementary Ed	Middle School Ed	High School Ed	Special Education	Other
Control (N=20)	20.0%	4.7%	4.7%	75.0%	23.8%
***PCI (N*=23)**	17.4%	4.30%	8.7%	52.2%	43.5%

Note. Teachers could select more than one category, so totals may exceed 100%.

Training

All *PCI* teachers received training in the implementation of the *PCI* program - 86% attended the original training, while 14% attended a make-up training at a later date. Only 29% of control teachers reported receiving any training for their current reading program, with only 14% percent of those teachers having been trained within the previous year.

As demonstrated in Figure 1and Figure 2, most *PCI* teachers who attended the full day training[6] felt the training for each of the specified content areas was crucial to being able to implement the program well. In addition, *PCI* teachers overwhelmingly agreed on the effectiveness of the training. Five of the six control teachers who received any training in their reading program reported it to be moderately to very effective, with the remaining teacher being of a neutral opinion.

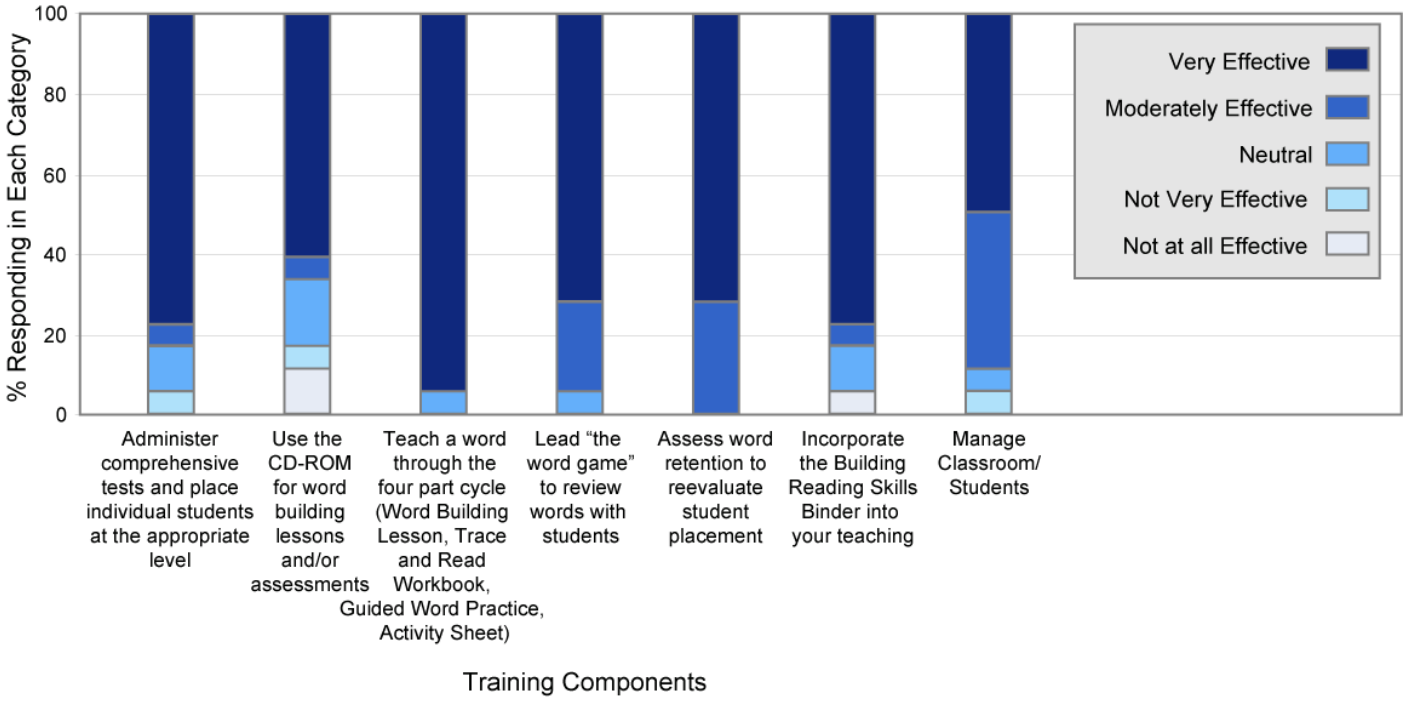


Figure 1. Effectiveness of *PCI* Training by Task Area

[6] Those who attended the makeup trainings were not asked for their opinions regarding training.

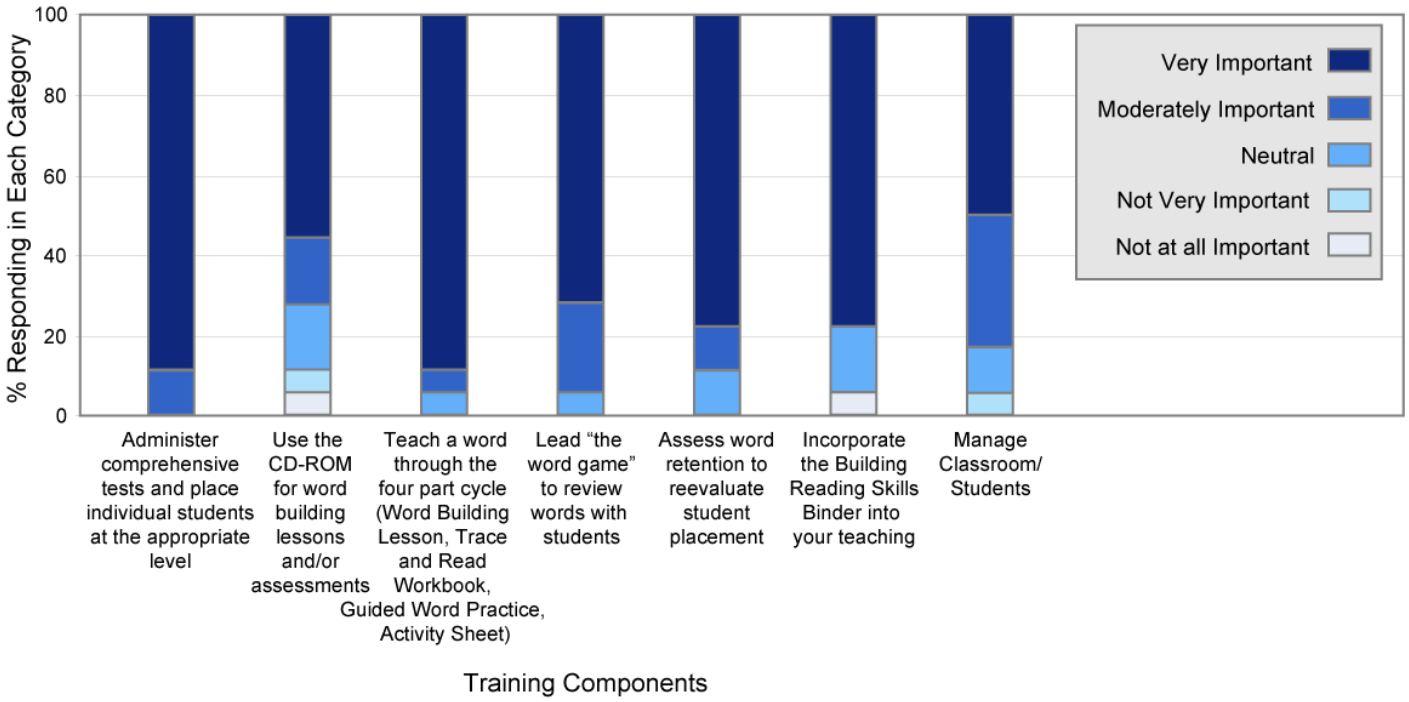


Figure 2. Importance of *PCI* Training by Task Area

Support

While there was no official follow-up training, *PCI* teachers were provided with contacts for both the program developers and the researchers, and were encouraged to seek support if necessary. By January 2008, only 3 teachers had reported contacting someone outside of their district for support. However, 50% of *PCI* teachers indicated seeking help from other teachers or district staff. Control teachers were not surveyed regarding support.

Availability of Materials

The 86% of the *PCI* teachers who attended the initial training in September and October received the *PCI* program at that time. By January 2008, 100% of *PCI* teachers reported to have all the materials they needed to fully implement the *PCI* program.

Availability of Teaching Assistants

At the beginning of the academic year, teachers indicated the best description for professionals who worked in their classroom with students participating in the *PCI* study. Among the 23 *PCI* teachers, all but three claimed to have additional assistance in their classrooms at the beginning of the year. Yet throughout the 7 surveys in which we asked teachers to account for the minutes of program instruction with all adults in the room, 8 (35%) never reported *PCI* instruction by another adult in the room. Therefore, while it appears that almost 90% of the teachers had another person who could help with *PCI* instruction, only 65% utilized this adult for PCI instruction. Although at the beginning of the school year 2 control teachers reported having no other adult to help with instruction, by the end of the 7 surveys every control teacher had reported another adult helping with instruction in their classroom. It is important to note, however, that the control teachers were reporting instruction for all language arts activities, while the *PCI* teachers were reporting only about their *PCI* instruction.

Table 16. Classroom Support for Reading Instruction

	Co-Teacher	Assistant for clerical tasks	Tutor for individual students	Aide or non-professional	Trained specialist for small groups	No professionals/ paraprofessionals	Other
Control (N=20)	10.0%	5.0%	5.0%	70.0%	10.0%	10.0%	20.0%
***PCI* (N=23)**	13.0%	4.3%	0.0%	69.6%	4.3%	13.0%	4.4%

Implementation in the Classroom

Here we describe how the programs were being implemented in participating classrooms. We will examine minutes of classroom instruction, how far *PC*I students progressed through the program, teacher fidelity to the prescribed program, and teacher opinions of the program and of student engagement.

Minutes of Instruction

By December 2007, only 77% of *PCI* teachers had begun instruction in the *PCI* program. However by January all but 1 teacher had begun instruction in the program. This non-implementing teacher never implemented the *PCI* program in her classroom during the 2007-08 academic year. Therefore, for the majority of teachers, *PCI* instruction began sometime in November and continued through to the end of the school year in May. A few teachers were able to continue teaching the program during summer school.

Table 17 shows that on average, *PCI* teachers reported completing the minimum daily instruction (20 minutes per day) required to meet the compliance standards set forth by the publisher. Only 3 teachers reported a daily average that approached the recommended 45 minutes of daily instruction.

Table 17. Minutes of *PCI* Instruction Per Day

Minimum	Maximum	Average	Median
2.4	84.2	26.4	20.2

Figure 3 shows the average minutes of *PCI* reading instruction over the course of the academic year. Implementation fluctuated between 400 and 500 minutes per week throughout the school year, except for a large dip in February, during which teachers and students in participating classrooms were involved with the Florida Alternate Assessment (FAA).

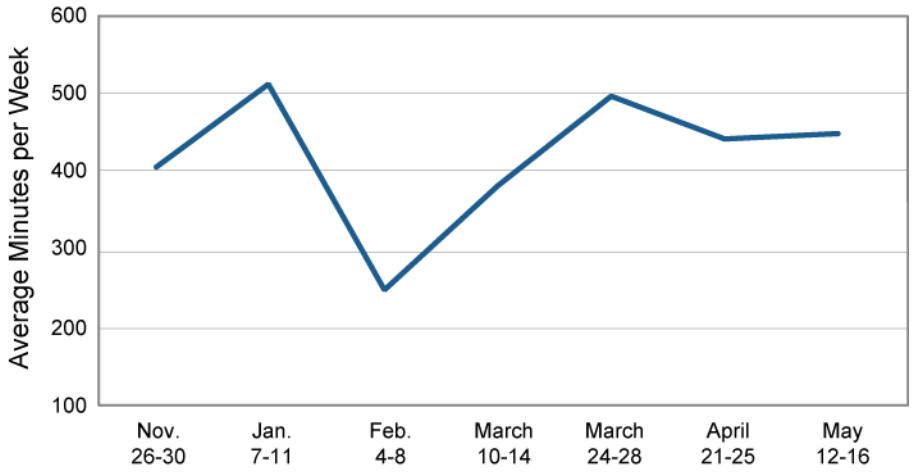


Figure 3. Average Weekly Minutes of *PCI* Instruction

While we did survey the control teachers about their instructional minutes, we learned that these numbers were not comparable to the minutes reported by the *PCI* teachers. For example, we asked the control teachers to report on the total minutes of reading instruction. However, we learned during classroom observations that *PCI* teachers were supplementing their instruction with other materials. So, in order to compare responses across the two groups, we would have needed to ask the *PCI* teachers about *total* minutes of reading instruction, rather than asking only about their minutes using *PCI*.

Student Progress

PCI expected the average student to complete the *PCI Reading Program - Level One* (master 140 words) within one school year, if receiving instruction in *PCI* for 45 minutes per day. However, our survey data, classroom observations, and end-of-the-year interviews all corroborated that the desired amount of instruction was not occurring. In fact, based on a sample of 50 *PCI* students at the end of the year, we found the average *PCI* student to have completed only 40 words during the course of the school year. Table 18 provides more detail on students' word completion.

Table 18. Number of Words Completed by May 2008

Minimum	Maximum	Average	Median
1 word	140 words	40 words	32 words

Fidelity to the Prescribed Program

Although much of our survey data show that *PCI* teachers were not supplementing the program with other materials, our classroom observations told a different story. In fact, as of April 2008 only 52% of *PCI* teachers claimed to have ever used any other materials to supplement instruction for students participating in the study. During the same month, we visited a sample of these classrooms and found that *PCI* was being supplemented with other materials in almost every case.

In 6 of the 8 *PCI* classes we observed, teachers utilized one-on-one instruction in teaching a new word to a student. Therefore only one student could receive instruction in *PCI* per adult in the room. When that student had completed their *PCI* lesson for the day, they would move to a

computer or work independently at their desk. This independent work was mostly spent with materials outside of the *PCI* program. The use of supplemental instructional materials in these classrooms can best be described as "a little bit of everything." Two of the classrooms we visited adapted the program to work for small group settings. In these cases, all students in the group were on the same word and the teacher had to review or modify instruction as necessary to keep the group moving along at the same pace. We did, however, find that *PCI* teachers were following the steps as specified in the teacher's guide.

The survey data in Table 19 through Table 21 show the extent to which teachers were following the prescribed lesson cycle. In most cases, the "NA" response was marked by the non-implementing teacher assigned to the *PCI* group. This question was asked 3 times over the course of implementation and the tables below display the averages of those responses.

Table 19. Which steps do you usually complete during each lesson cycle?

	Always	Sometimes	Never	NA
Step 1a. Learn a Word: Word Building Lesson	86.0%	10.5%	0.0%	3.5%
Step 1b. Learn a Word: Trace and Read Workbook	80.7%	15.8%	0.0%	3.5%
Step 1c. Learn a Word: Guided Word Practice	80.7%	15.8%	0.0%	3.5%
Step 1d. Learn a Word: Activity Sheets	61.4%	33.3%	0.0%	5.3%
Step 2. Repeat steps 1a-1d to learn four new words	63.2%	29.8%	0.0%	7.0%
Step 3. Review the words with "The Word Game"	33.3%	47.4%	8.8%	10.5%
Step 4. Assess Word Retention	63.2%	22.8%	3.5%	10.5%
Step 5. Read a Book	80.7%	12.3%	0.0%	7.0%

Table 20. How do you usually organize students during this step?

	One-on-one	Group instruction	Independent work	NA
Step 1a. Learn a Word: Word Building Lesson	82.5%	14.0%	0.0%	3.5%
Step 1b. Learn a Word: Trace and Read Workbook	63.2%	24.6%	8.8%	3.5%
Step 1c. Learn a Word: Guided Word Practice	78.9%	17.5%	0.0%	3.5%
Step 1d. Learn a Word: Activity Sheets	35.1%	22.8%	38.6%	3.5%
Step 2. Repeat steps 1a-1d to learn four new words	71.9%	22.8%	0.0%	5.3%
Step 3. Review the words with "The Word Game"	22.8%	61.4%	1.8%	14.0%
Step 4. Assess Word Retention	78.9%	10.5%	1.8%	8.8%
Step 5. Read a Book	78.9%	14.0%	1.8%	5.3%

Table 21. Who is this step usually taught by?

	Teacher	Other adult	NA
Step 1a. Learn a Word: Word Building Lesson	96.5%	0.0%	3.5%
Step 1b. Learn a Word: Trace and Read Workbook	82.5%	14.0%	3.5%
Step 1c. Learn a Word: Guided Word Practice	94.7%	1.8%	3.5%
Step 1d. Learn a Word: Activity Sheets	63.2%	28.1%	8.8%
Step 2. Repeat steps 1a-1d to learn four new words	91.2%	5.3%	3.5%
Step 3. Review the words with "The Word Game"	61.4%	19.3%	19.3%
Step 4. Assess Word Retention	86.0%	1.8%	12.3%
Step 5. Read a Book	84.2%	8.8%	7.0%

The Activity Sheets are part of the mandatory lesson cycle but because they are reproducible, teachers have the *option* of sending them as homework. As of late April, 81% of teachers reported using the resource for in-class exercises, 38% of the teachers sent activity sheets as homework, and 20% (2 teachers) reported never having used the resource.

In late April, when asked whether they used print materials or the CD-ROM for word building lessons and/or assessments, 5% of teachers reported using only the CD-ROM for these tasks, 61% reported using only the print materials, and 29% of *PCI* teachers used both the CD-ROM and print materials at some point. Only the non-implementing teacher reported never having used either.

An optional supplement provided by the program to address students with additional needs exists in the form of the Building Reading Skills Binder. This resource is available for teachers who have students that may need additional support, including help with phonics. Although not a mandatory part of the program, by late April, 43% of teachers had utilized this resource.

PCI teachers also made adaptations to the materials based on need. We found teachers who utilized the materials to make games, worksheets, and various other activities to help solidify learning.

Teacher Opinions

During the training sessions for the pre and post assessments, both control and *PCI* teachers expressed the need for a reading program targeted toward their student population that was engaging for students. At the end of the academic year, we asked teachers in both assignment groups for opinions regarding their current curriculum. For control teachers, because of the extreme variation among teachers as far as materials used, we were only able to ask these questions in general terms. As seen in Table 22 and Figure 4, 65% of *PCI* teachers reported that they were very satisfied with their reading program, while only 21% of control teacher reported that opinion. Not one *PCI* teacher reported dissatisfaction or even a neutral opinion of the program.

Table 22. Teacher opinion of Reading Program: Control Versus *PCI*

Materials	Very satisfied	Somewhat satisfied	Neutral	Somewhat dissatisfied	Very dissatisfied	Haven't used enough
Control (N=19)	21.0%	42.0%	21.0%	0.0%	11.0%	5.0%
***PCI* (N=20)**	65.0%	30.0%	0.0%	0.0%	0.0%	5.0%

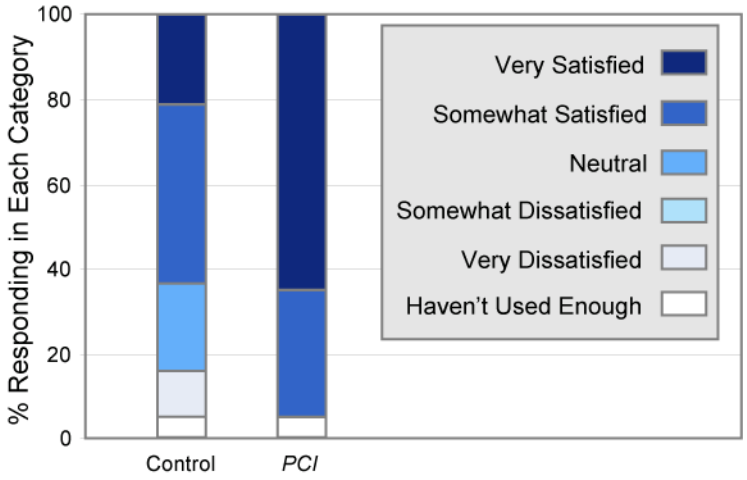


Figure 4 Teacher Opinion of Reading Program: Control Versus *PCI*

In addition, we also asked teachers in both groups whether they would recommend their reading program to other teachers. 95% of *PCI* teachers claimed that they would recommend the program, with the remaining 5% representing the one non-compliant teacher who had not used the program enough to form an opinion. Only 63% of control teachers would recommend their current program to teachers of this population, with 26% of the teachers actually stating that they would not recommend the program, and the remainder not having an opinion.

Table 23 elaborates on how teachers felt about the various optional or supplemental components of the *PCI* program. For each component, teachers who used each piece of the program enough to form an opinion were generally satisfied with the various aspects of the program.

Table 23. Teacher Satisfaction with Optional *PCI* Components

Materials	Very satisfied	Somewhat satisfied	Neutral	Somewhat dissatisfied	Very dissatisfied	Haven't used enough
***PCI* Building Reading Skills Binder**	40.0%	20.0%	20.0%	0.0%	0.0%	20.0%
***PCI* CD-ROM**	30.0%	15.0%	0.0%	10.0%	5.0%	40.0%

Overall, *PCI* teachers expressed satisfaction with the program. They also expressed a desire to keep using the program. In the final survey of the year, we asked those currently assigned to the *PCI* group if they planned to continue using the program once the research study was over (following the 2007-08 academic year). 85% of participating *PCI* teachers planned to use the program again the following school year and only one teacher intended to discontinue teaching the program.

Table 24. Do You Believe You Will Continue Teaching the *PCI Reading Program - Level One* Once This Research Study Is Complete?

Yes, I plan to increase use	Yes, I plan to continue	Yes, but I plan to decrease use	No, I don't plan to continue	I don't know
65.0%	20.0%	0.0%	5.0%	10.0%

In addition to asking about the teachers' own opinions, an earlier survey asked teachers about student engagement and enjoyment of the program. Teachers were reminded to capture the feelings of their students, as opposed to how they themselves might feel about the program. When asked how they would rate student enjoyment of learning reading through the *PCI* program, teachers responded very positively. While control teachers mostly reported moderate to high levels of student enjoyment, 73% of *PCI* teachers reported high to very high levels of student enjoyment.

Table 25. Student Enjoyment

	Very high	High	Moderate	Low	Very low	I don't know
Control (N=20)	5.0%	40.0%	45.0%	0.0%	0.0%	10.0%
***PCI* (N=22)**	31.8%	40.9%	18.2%	4.5%	0.0%	4.5%

Teachers were also asked to rate student level of engagement while participating in various aspects of the program. Students would be considered fully engaged if they displayed consistent on-task behavior.

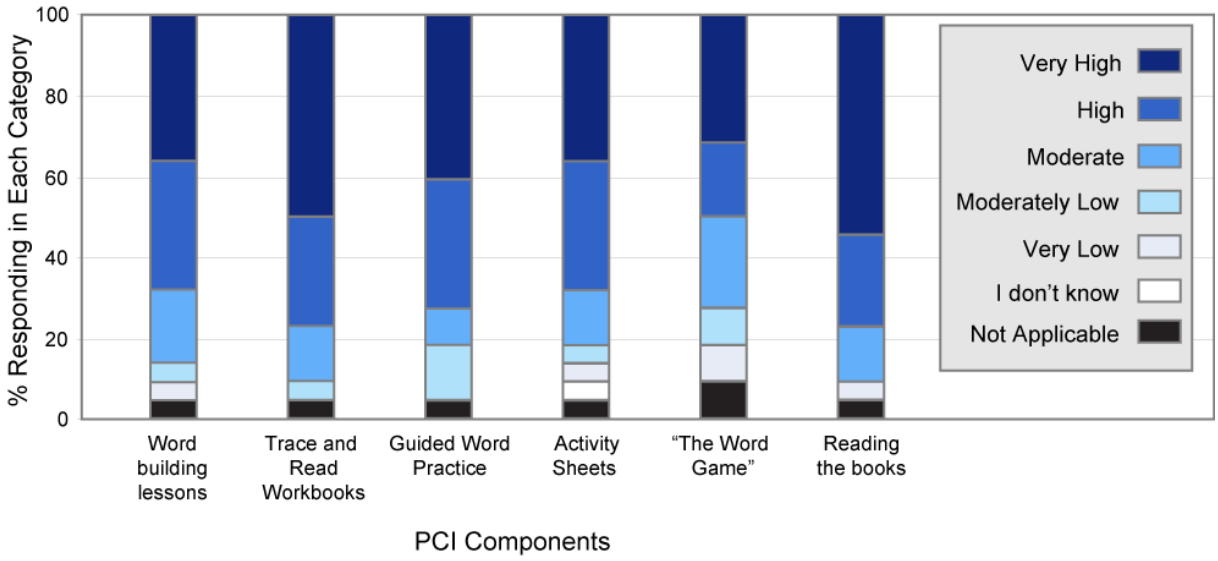


Figure 5. Levels of Student Engagement With *PCI* Components

As seen, teachers observed high levels of engagement and enjoyment in their students overall. Certain activities, such as reading books and the Trace and Read Workbooks ranked very high in terms of student engagement, while other activities such as the Word Game received more moderate rankings.

Additionally, teachers provided many descriptive comments about the program when asked for their opinions in an open-ended format. From these responses we discovered common themes with regard to the successes and challenges of implementation. When asked to describe aspects of the program that they liked, teachers commonly mentioned the variety of materials available in the program. Teachers connected the materials to high levels of student engagement. Another frequently reported success among the free response questions was that students were actually retaining words and learning to read books. One teacher expresses a common sentiment: "Their faces light up and their whole week is made when they are able to read another book." The challenges teachers reported mostly reflected issues of classroom management. Many teachers, especially those with no or little classroom support, found it difficult to make time for individualized *PCI* instruction while managing the rest of their students. Overall these teachers believed that with more time or classroom support, the program would work very well with their students.

Summary of Implementation

The conditions for implementation of the *PCI* program appeared to be quite good. All teachers in the *PCI* program received training, while few teachers in the control group had ever received any training in their reading program. Eighty-six percent of the *PCI* teachers had all the materials they needed at the time of the training. The *PCI* teachers were offered support from PCI although no teachers appeared to take advantage of this option. However participants did go to each other, or district supervisors, for help.

PCI teachers appeared to fully implement the program, although at minimum rather than optimal levels. Through the course of the academic year, teachers in the *PCI* group recorded accomplishing the minimum amount of instruction as specified by the publisher for implementation compliance. Most all of these teachers supplemented the *PCI* program with other curricular materials. While students did not progress as far in the program as initially expected, teachers did generally follow the lesson cycle as specified by the publisher. Teachers in the *PCI* program reported higher levels of student enjoyment, as well as general satisfaction with the program, than did teachers in the control group.

Student-level Impact Results

In this section we turn to the examination of the results at the student level where the outcome measure was the sight word test. We look at the impact of *PCI* on this outcome as well as examine the moderating effects of variables we identified in our research questions: our pretests (sight word and phonological), grade and teacher experience.

We first explain how we subdivided the students depending on their pretest performance. This is important for the clarity of the results since we identify what we consider a qualitative difference among the students in the understanding of the sight word recognition task.

Summary of Sample Subdivision

After limiting cases as described in the previous section, our sample consists of 128 students (61 in the *PCI* condition and 67 in the control condition) and 35 teachers (20 *PCI* and 15 control). However, we decided to further divide the sample according to scores received on pretest and posttest. Table 26 summarizes the breakdown of students and teachers in each category.

Table 26. Breakdown of Students Who Qualify for the Final Analysis

	Control		*PCI*		
	No. of students	No. of teachers with at least one student meeting criteria	No. of students	No. of teachers with at least one student meeting criteria	Total number of students
Pretest score = 0, posttest = 0	9	6	0	0	**9**
Pretest score = 0, posttest = at least 1	13	8	16	8	**29**
Pretest and posttest score = at least 1	38	11	51	16	**89**
Pretest score = at least 1, posttest = 0	1	1	0	0	**1**
Totals	**61**		**67**		**128**

We separated this sample into two groups and ran separate analyses on each. We considered separately students who had a zero on the pretest and those who score above zero on the pretest, for two reasons: First, a pretest score of zero curtailed the distribution of scores along the horizontal (pretest) axis. We did not want the distribution of points along the y-axis to drive the estimates based on the scatterplot overall. For example, we didn't want a concentration of points at the origin (i.e., representing scores for students who obtained zero correct at pretest and posttest) to determine the interaction between the pretest and treatment for students who received more than one word right on the pretest. Second (a point related to the first) we believed that students who received a score of zero on the pretest potentially represented a different kind of student than those who scored above zero on the pretest. It is possible that these students did not understand or were not engaged in the task of identifying words.

For these reasons we decided to analyze the results for these mutually exclusive groups of students separately. For those who scored one or higher on the pretest as well as the posttest (89 students), we can be sure that the students understood the tests. For these students we analyze the full range of results. For the verbal students who score zero on the pretest (38 students) we

compare the difference between the *PCI* and control group in the proportion who score above zero on the posttest as well as in the average number words read correctly.

Results for Verbal Students who Score Zero on the Pretest

We begin with the results for students who received a zero score on the Sight Word pretest. Table 27 shows the difference between the *PCI* and control conditions in the number of students who score higher than zero on the posttest. We observe that all the students in the *PCI* condition who started at zero, score above zero on the posttest, whereas only some of the students in the control condition who start at zero get above zero on the posttest. A very low *p* value, which we obtained by running Fisher's exact test, shows that this is probably not a chance result.

We were concerned that this strong result reflected differences between *PCI* and control teachers in the extent to which they tested certain kinds of students – specifically, that *PCI* teachers were more likely to not retest students who they believed did not stand a fair chance of benefiting from the program. This would have been one explanation for there being no students in the *PCI* condition who scored zero on the pretest as well as on the posttest. But, if students who scored zero on the pretest were selectively not retested in the *PCI* condition we would have expected these students to inflate the counts of student scoring zero on the pretest and having no score on the posttest in this condition. However, the number of students fitting this profile was 15 and 10 in the control and *PCI* conditions, respectively. With these counts we don't believe that students who scored zero on the pretest were selectively excluded from the *PCI* condition prior to posttest, and that the difference between conditions in the number of cases who score zero on the pretest and above zero on the posttest is due to the effectiveness of *PCI.*

Table 27. Comparison of Students Having a Score of Zero on Pretest Between *PCI* and Control Group

	Posttest is zero		
Condition	**Yes**	**No**	**Totals**
Control	9	13	**22**
PCI	0	16	**16**
Totals	**9**	**29**	**38**
Statistics		**Value**	***p* value**
Fisher's exact test		0.00	<.01

We also tested the mean difference in sight word posttest performance between students in the *PCI* and control conditions for the 29 students who started with a zero pretest and who scored higher than zero on the posttest. The results are shown in Table 28 below. The *p* value for the estimate of the mean difference is .02, which gives us a high level of confidence that there is a true difference.

For this result we considered whether teachers in the *PCI* condition were more likely to not retest students who score low on the pretest, thereby leading to an inflated difference between conditions on the posttest. If this were the case, we would expect to see a disproportionate number of students not being retested in the *PCI* condition (which, as noted above, we did not observe) or we would expect to see a larger proportion of students in *PCI* who are either missing the pretest or who are not on the roster (i.e., if teachers did not submit results for students after observing their pretest scores.) Neither of these patterns is observed which gives us a stronger warrant for concluding that the difference we observe is due to the effect of exposure to *PCI*[7].

Table 28. Differences in Sight Word Posttest Scores for Students Who Score Zero on the Pretest and at Least One Correct on the Posttest in *PCI* and Control Groups

Descriptive statistics: Sight Word pretest	Raw group means	Standard deviation	Number of students	# of teachers	Effect size
Control	1.85	2.48	13	8	1.13
PCI	7.00	5.53	16	8	
Fixed effects	**Estimate**	**Standard error**	**DF**	***t* value**	***p* value**
Control	1.82	1.30	14	1.40	.18
Effect of *PCI*	4.94	1.79	14	2.76	.02
Random effects[a]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	2.95	3.98		0.74	.23
Within-teacher variation	16.71	5.41		3.10	<.01

[a] Teachers were modeled as a random factor.

Different estimation approaches lead to small differences in estimates of the same parameters in this table.

Results for Verbal Students who Score Higher than Zero on the Pretest

Next we address *PCI* outcomes using the Sight Word assessment scale for students who received a score of one or higher on the pretest. Table 29 provides a summary of the sample we used and the results for the comparison of Sight Word assessment scores for students in *PCI* and control groups. The "Unadjusted" row gives information about all the students in the original sample for whom we have a pretest and posttest. This shows the means and standard deviations as well as counts for students, teachers and schools in that group. The last two columns provide the effect size, that is, the size of the difference between the means for *PCI* and control groups in standard

[7] This impact estimate is conditional on students scoring above zero on the posttest. If we model proficiency as distributed along a latent scale, and if we assume that students have to cross a threshold to score above zero, then the result shown here is computed for students above the threshold, where we assume that students in the *PCI* condition have all passed the threshold, but where some students in the control condition lie below the threshold and are not included. Presumably, including students who, at the time of the posttest, are below the threshold into the impact estimate would increase the effect size by lowering the average performance for the control group.

deviation units. Also provided is the *p* value, indicating the probability of arriving at a difference as large as, or larger than, the absolute value of the one observed when there truly is no difference. The "Adjusted" row is based on the same sample of students. The mean difference, and therefore the effect size, is adjusted to take into account the student pretest scores; hence, these estimates are adjusted for any chance imbalances on the pretest between the two randomized groups.

Table 29. Effect Sizes for Students with Posttests and Who Score Higher than Zero on the Pretest

	Condition	Means	Standard deviations[a]	No. of students	No. of teachers	No. of schools	Effect size	*p* value[b]	Percentile standing
Un-adjusted effect size	**Control**	8.50	6.19	38	11	11	.59	<.01	22%
	PCI	11.94	5.37	51	13	16			
Adjusted effect size	**Control**	8.50	Same sample used.				.55	<.01	21%
	PCI	11.67							

[a] The standard deviations used to calculate the adjusted and unadjusted effect sizes are calculated from the scores of the students in the sample for that row.

[b] The unadjusted effect size is Hedges' g adjusted for clustering of students in teachers (Hedges, 2007). The adjusted effect size is based on the impact estimate obtained from SAS PROC MIXED where we model clustering of students in teachers (i.e., a two-level random intercept model.) We also include pretest modeled at the student level. The estimate of the standard deviation that is used in the denominator of the adjusted effect size calculation is the same as the one used to compute Hedges' g. The *p* value is for the estimate of the treatment effect for this model.

Figure 6 provides a visual representation of the information in Table 29. The bar graphs represent average performance using the metric of the Sight Word assessment.

The panel on the left shows average pre- and posttest scores for the control and *PCI* groups. The pre- and posttest bars show that, on average, both the *PCI* and control groups grew in their sight word reading achievement during the year, although the control grew very little in comparison.

The panel on the right is a visual display of results from the row labeled "Adjusted" in Table 29. It shows estimated performance on the posttest for the two groups based on a statistical equation that adjusts for students' pretest scores. The overall reading effect size (in standard deviation units) is .55, which is equivalent to a gain of 21 percentile points for the median student if s/he had received *PCI* instead of being in the control condition. The low *p* value for the *PCI* effect (<.01) indicates we should have high confidence that the actual difference is different from zero. We added 80% confidence intervals to the tops of the bars in the figure. The lack of overlap in these intervals further indicates that we are not just seeing a chance difference.

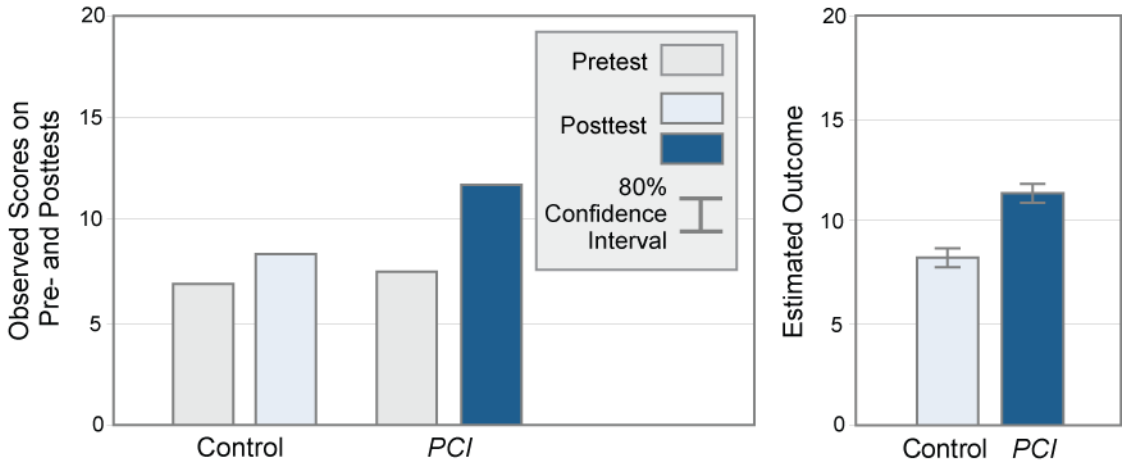


Figure 6. Impact on Sight Word Assessment: Unadjusted Pre- and Posttest Means for Control and *PCI* (Left); Adjusted Means for Control and *PCI* (Right)

Moderating Variables

We now report on our examination of the moderating effects of other variables (Sight Word Pre-Assessment, Phonological Pre-Assessment, grade level and years of experience teaching special education). We provide a separate table of results for most of these moderator analyses. The fixed factor part of each table provides estimates of the factors of interest.

Including Sight Word Pretest as a Moderator

We first show whether the impact of *PCI* is different for students at different levels of prior achievement on the sight word pretest. At the bottom of the table we give results for technical review—these often consist of what are called random effects estimates. As was described earlier in this report, random effects are added to the statistical equation to account for dependencies in observed scores that happen because students come from the same classes or teachers.

Table 30 shows the estimated impact of *PCI* on the performance of students with an average pretest in reading as measured by the Sight Word assessment as well as the moderating effect of their prior scores.

Table 30. Moderating Effect of the Sight Word Pretest on the Impact of *PCI* on Student Performance on the Sight Word Assessment

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for a control student with an average pretest	8.92	0.59	24	15.13	<.01
Change in outcome for a control student for each additional word correct on the pretest	0.92	0.10	60	9.41	<.01
Effect of *PCI* for a student with an average pretest	2.87	0.77	24	3.73	<.01
Change in the effect of *PCI* for each additional word correct on the pretest	-0.15	0.13	60	-1.17	.25
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	0.29	1.02		0.28	.39
Within-teacher variation	11.22	1.94		5.79	<.01

[a] Pair fixed effects accounted for a trivial amount of variance on the posttest and are not modeled.

[b] The prior score was centered at its mean.

[c] The teacher-level slope for pretest was first modeled to vary randomly so that it reflects sampling variation at the teacher level, which is consistent with the way we model teacher-level intercepts. The variance component for this effect was trivially different from zero. When this happens, SAS PROC MIXED automatically constrains the parameter estimate be zero (a point that we confirmed through consultation with SAS.) We do not display this result in the table above.

The row in the table labeled "Effect of *PCI* for a student with an average pretest" tells us whether *PCI* made a difference on Sight Word assessment for a student who has an average score on the pretest. The estimate associated with *PCI* is 2.87. This shows a positive effect of *PCI* for the student with an average pretest. The *p* value of <.01 indicates that this result is unlikely to be due to chance. Using the criteria outlined earlier in the report, we conclude that we have a high level of confidence that the true impact is different from zero.

We also estimated the moderating effect of the pretest score on the impact of *PCI* to determine whether the intervention was differentially effective for students along the pretest scale. The coefficient for the interaction between sight word pretest and the treatment condition is -0.15, which shows a small decrease in the *PCI* effect with each one-unit increase on the pretest. The *p* value of .25 gives us no confidence that the true differential impact is different from zero.

As a visual representation of the results described in Table 30, we present a scatterplot in Figure 7, which graphs student growth over the school year in terms of reading achievement as measured by Sight Word assessment scores. This graph shows where each student started in terms of his or her pretest score (horizontal x-axis) and his or her outcome score (vertical y-

axis). Each point plots one student's post-intervention score against his or her pre-intervention score. The darker points represent *PCI* students; the lighter points, control students. The shaded area in the lower right of the graph is the area of negative change (i.e., where students lost ground).

The two lines are the estimated values on the posttest for students in the *PCI* and control conditions. Consistent with the results described above, we observe that the average performance for the *PCI* students is higher than for the control students and that the effect diminishes slightly with higher sight word pretest scores. However, as noted above, we have no confidence that this differential effect is due to factors other than chance.

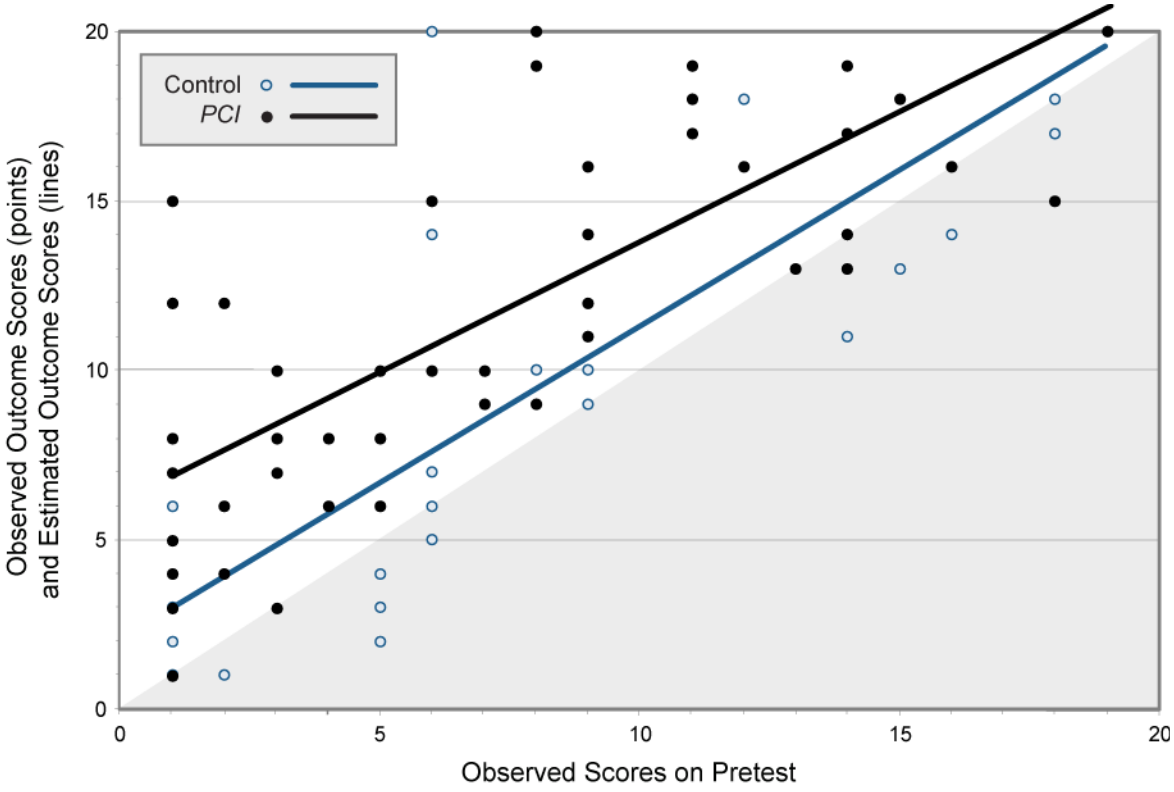


Figure 7. Comparison of Estimated and Actual Outcomes for *PCI* and Control Group Students (Sight Word Achievement)

Including Phonological Pretest as a Moderator

Next, we examine whether the impact of *PCI* is different for students with different pretest scores on a test of phonological skills. We use what are called residualized scores[8]. This is important because phonological skills and sight word skills may be correlated and we want to

[8] To obtain this score we regressed pretest scores on a test of phonological skills against the sight word achievement pretest and used the residuals from that regression as the moderator. The residuals are a measure of how well students perform on that part of their measured phonological skills that is not predicted from their sight word performance.

know about the effect of the student's phonological skills independent of the effect of their sight word skills. The residualized scores allows us to assess whether phonological skills, 'net of' sight word skills, moderate the impact of *PCI* on sight word reading skills[9]. Table 31 shows the results.

Table 31. Moderating Effect of Phonological Skills on the Impact of *PCI* on Sight Word Performance

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for a control student with an average sight word pretest and an average PH pretest	8.33	0.53	25	15.68	<.01
Change in control outcome for each additional word correct on the sight word pretest	0.82	0.06	58	14.54	<.01
Change in control outcome for each unit-increase on the PH pretest	-0.15	0.17	58	-0.86	.39
Effect of *PCI* for a student with an average sight word pretest and an average PH pretest	3.49	0.69	25	5.04	<.01
Change in effect of *PCI* for each unit-increase in the PH pretest	0.35	0.21	58	1.64	.11
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Teacher mean achievement	0.26	0.90		0.28	.39
Within-teacher variation	8.16	1.66		4.93	<.01

[a] Pair fixed effects accounted for a trivial amount of variance on the posttest and are not modeled.

[b] Both the sight word pretest and the PH pretest are centered on their respective means.

[c] The teacher-level slope for the PH pretest was first modeled to vary randomly so that it reflects sampling variation at the teacher level, which is consistent with the way we model teacher-level intercepts. The variance component for this effect was trivially different from zero. When this happens, SAS PROC MIXED automatically constrains the parameter estimate be zero (a point that we confirmed through consultation with SAS.) We do not display this result in the table above.

[9] From here forward, when we refer to phonological skills or the PH pretest, we mean a measure of those skills based on residualized scores.

The row in the table labeled “Effect of *PCI* for a student with an average sight word pretest and an average PH pretest” tells us whether *PCI* made a difference on sight word achievement for a student with average pretest scores. The estimate associated with *PCI* is 3.49. This shows a positive impact of *PCI*. The *p* value of <.01 indicates that we can have a high level of confidence that the true impact is different from zero.

We also estimated the moderating effect of the PH pretest score on the impact of *PCI* to determine whether the intervention was differentially effective for students at different points along the PH pretest scale. The coefficient associated with the interaction of the PH pretest with *PCI* is 0.35, which shows a small increase in the *PCI* effect with each one-unit increase on the PH pretest. The *p* value of .11 indicates that we can have some confidence that the true differential impact is different from zero. In other words, that the effect of *PCI* is slightly more effective for students who have higher scores on the pretest of phonological skills.

Figure 8 displays the same information represented in Table 31 but this time in the form of a bar graph showing the estimated difference between *PCI* and control conditions for students at the medians of the first and fourth quartiles of the PH pretest and who have an average score on the sight word pretest. The bar graph includes the 80% confidence interval as a marker at the top of the bars. We see that, for a student both at the median of the first quartile and at the median of the fourth quartile, there is no overlap in the confidence intervals. This demonstrates the strong average effect of *PCI* that we described earlier. In addition, we see a bigger difference in performance between students in the two conditions at the median of the fourth quartile than at the median of the first quartile. This is a graphic display of the interaction discussed above, where we have some confidence that *PCI* has a bigger impact at the high end of the PH pretest reading scale.[10] In other words it appears that *PCI* provides greater benefit for students who enter with a higher level of phonological skills.

[10] The standard errors and significance levels for the estimate of the moderator effect in the table of results is not exactly the same as those used to express uncertainty in the graphs. The standard error for the interaction expresses uncertainty in the parameter that measures a difference in impact (i.e., the difference in the difference between treatment and control for each unit-increase in the moderator). The standard errors used for the confidence intervals in the graphs express uncertainty in the impact (i.e., the difference between treatment and control) at the given level of the moderator.

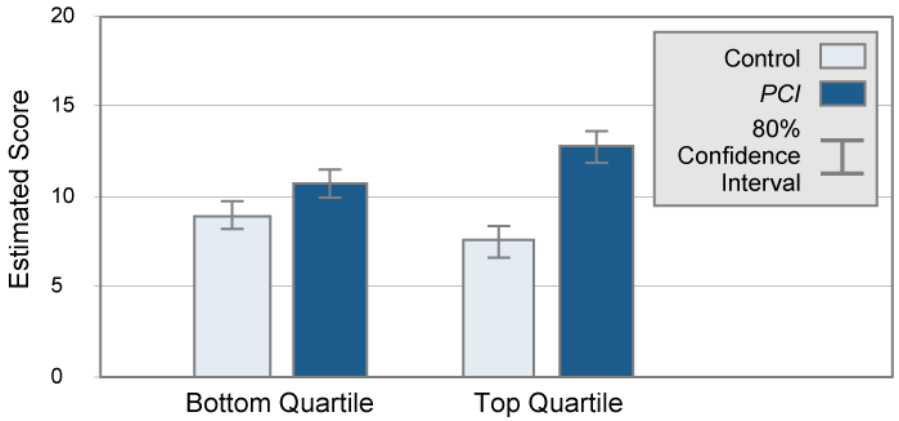


Figure 8. For Sight Word Outcomes: Impacts for Students Who are at the Medians of the Top and Bottom Quartiles of the Pretest of Phonological Skills

Including Grade Level as a Moderator

We also considered whether *PCI* is differentially effective for students in different grades. Table 32 shows the moderating effect of grade level on the impact of *PCI* on students' performance on the Sight Word assessment. We treated grade level as a continuous variable for this analysis.

We first consider the row in the table labeled "Effect of *PCI* for a student with an average pretest and who is in Grade 6." The estimate of the effect of *PCI* is 3.37. This shows a positive impact of *PCI* for this type of student. The *p* value of <.01 indicates that we can have a high level of confidence that the true impact is different from zero.

We also estimated the moderating effect of grade on the impact of *PCI* to determine whether the intervention was differentially effective for students at different grades. The coefficient associated with the interaction of grade with *PCI* is .27 which shows a small increase in the *PCI* effect with each additional grade. The *p* value of .52 indicates that we have no confidence that the true differential impact is different from zero.

Table 32. Moderating Effect of Grade on the Effect of *PCI* on Sight Word Performance

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
Outcome for a control student with an average pretest and who is in Grade 6	8.54	0.51	19	16.67	<.01
Change in outcome for each unit-increase on the pretest	0.88	0.06	62	14.62	<.01
Change in control outcome for each one-grade-increase	0.31	0.33	62	0.92	.36
Effect of *PCI* for a student with an average pretest and who is in Grade 6	3.37	0.66	62	5.07	<.01
Change in effect of *PCI* for each one-grade-increase	0.27	0.41	62	0.65	.52
Random effects[b]	**Estimate**	**Standard error**		***z* value**	***p* value**
Assignment pair mean achievement	0.04	0.70		0.06	.48
Within-teacher variation	9.02	1.54		5.87	<.01

[a] Grade is centered on grade 6.

[b] It was necessary to model a random intercept for pairs instead of teachers for the model to converge.

The teacher-level slope for grade level was first modeled to vary randomly so that it reflects sampling variation at the teacher level, which is consistent with the way we model teacher-level intercepts. The variance component for this effect was trivially different from zero. When this happens, SAS PROC MIXED automatically constrains the parameter estimate be zero (a point that we confirmed through consultation with SAS.) We do not display this result in the table above.

As an additional exploration related to these results, we examined the correlation between grade level and pretest score. While the above results do not indicate a difference in outcome for the control group depending on grade, this assumes that the pretest is held constant; in other words, it shows the association between the posttest and grade-level net of pretest. However, when we examined the Pearson's correlation between grade level and pretest scores directly, we found a very low correlation of 0.11 with a *p* value of .17. This low correlation in sight word recognition and grade across a span of at least four years suggests that little reading skill is being acquired in the programs for this population in the years prior to the experiment.

Including Special Education Teaching Experience as a Moderator

Although we had planned to look at whether students of experienced or inexperienced teachers benefit more from *PCI*, there were not a sufficient number of inexperienced teachers (only six altogether) to make a statistical comparison.

Discussion

In this randomized experiment, we investigated whether students whose teachers were given the *PCI Reading Program - Level One* (*PCI*) achieved higher sight word assessment scores in reading than students whose teachers did not have the program. We also investigated whether *PCI* had a different effect for specific subgroups of students: those who scored low on the sight word and phonological pre-assessments, those in lower grades, and those whose teachers had more experience teaching special education. Our sample was composed of students with severe disabilities, and their teachers, from two Florida districts: Brevard Public Schools and Miami-Dade County Public Schools. Our outcome measure was a sight word assessment developed by an independent consultant.

We approached this experiment as an efficacy trial in certain respects. In particular, as a new program being tested for the first time with a challenging population of students, we wanted to know whether it could achieve its intended purpose: teaching specific sight words. Our pre- and posttest consisted of a sample of words taken from the *PCI* program itself. It was not a general test of reading but rather one that was closely aligned to the program. The results of this experiment showed that students in the *PCI* program had substantially greater success in learning sight words than students in the control group—a difference equivalent to a 21 percentile point improvement.

We must note that the experiment suffered substantial attrition of students as well as teachers. The major reason for loss of students from the experiment was the difficulty in obtaining parental consent. Despite the loss of students, the experiment was still able to detect a very large impact on the sight word outcome measure.

We also had to address a problem inherent in a sight word test. Many students in our sample were unable to name any words on the pretest. Our solution was to conduct two separate tests. With the sample of students who remained following attrition, we considered separately students who had a zero score on the pretest and those who scored above zero on the pretest, as we believed the composition of these groups could differ fundamentally. For example, it is possible that students who had a zero score on the pretest did not understand or were not engaged in the task of identifying words. Non-verbal students were excluded from this study.

Within the group of students who scored zero on the pre-assessment, we have a high level of confidence that there is a positive difference in outcomes between control and *PCI* student groups. Students in *PCI* classrooms performed significantly higher than students in control classrooms. For the group of students that scored at least one point on the pre-assessment, we again found a significant impact on the *PCI* treatment. Both unadjusted and adjusted analyses show high effect sizes (.55, .59) with small *p* values.

In examining moderating variables, we found the sight word pre-assessment to not be significant in determining the impact on student outcomes on the post-assessment. However, we did find the Phonological Assessment to have a small moderating effect. We can have some confidence that students with higher scores on the Phonological Assessment benefit more from *PCI* than students who scored lower. On the other hand, with respect to grade level, we see no significant effects. That is, the effect of *PCI* is not influenced by the grade level of the student. For example, 8th graders are not benefitting more from *PCI* than 4th graders. Also, because the small number of teachers with fewer than four years experience teaching Special Education prevented us from carrying out a statistical comparison, we were unable to examine whether years of teaching Special Education moderated the effect of *PCI.*

Although *PCI* teachers had all the materials necessary to implement the program early on, we found that teachers were completing the minimum amount of *PCI* instruction specified by the publisher. While we know that nearly all *PCI* teachers were supplementing *PCI* instruction with other reading materials, we do not know how much time they spent doing so. For this reason, we were unable to compare minutes of reading instruction across the control and *PCI* groups. While *PCI* teachers did spend less time implementing the program than expected, they had very positive opinions of the program and, for the most part, planned to continue implementation the following year.

Our research from the first year of the study shows that it takes longer for one student to complete the *PCI Reading Program - Level One* than initially estimated by the publishers. In addition, the fact that grade level did not influence outcomes shows a need for more research within this population of students. We are continuing our research of the *PCI Reading Program* in both districts over the next four years in order to follow students through the *PCI Reading Program - Level Two* and into *Level Three*.

Our initial experimental study provides evidence of the efficacy of the *PCI Reading Program*. The positive results for students and positive acceptance by the teachers is useful information for school districts looking for a reading program for severely disabled students. We consider our results preliminary because we tested only *Level One* and our achievement measure was limited to the specific goals of the program. Still, these results are encouraging and call for continued and broader investigation of this promising program.

References

Amrein-Beardsley, A. L. (2006). Teacher research informing policy: An analysis of research on highly qualified teaching and NCLB. *Essays in Education*, 17

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs For Generalized Causal Inference.* Boston, MA: Houghton Mifflin.

The Center for Public Education (2005, October 4) Teacher quality and student achievement: Key lessons learned. Retrieved January 9, 2008, from http://www.centerforpubliceducation.org/site/c.kjJXJ5MPIwE/ b.1510943/k.9EF1/Teacher_quality_and_student_achievement_ Key_lessons_learned.htm