RESEARCH REPORT

Comparative Effectiveness of PCI Education's *PCI Reading Program:* Phase 2

A Report of a Comparison Group Study in Brevard Public Schools and Miami-Dade County Public Schools

Jenna Zacamy
Xiaohui Zheng
Andrew Jaciw
Boya Ma
Qingfeng Zhao

Empirical Education Inc.

March 2010

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.

The methodology used in this report was presented as a poster session at the Society for Research on Educational Effectiveness Annual Conference in Washington D.C., March 2010.

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 comparative effectiveness of the *PCI Reading Program* through a five-year longitudinal study. Phase 1 of the study consisted of a randomized control trial studying the efficacy of the *PCI Reading Program- Level One* that was conducted in the 2007-2008 in Miami-Dade County Public Schools and Brevard Public Schools with supported level students in grades 3–8 and their teachers. This report presents the findings from Phase 2, which studied the efficacy of the *PCI Reading Program- Level One* and *Level Two*. Phase 2 was built upon the Phase 1 RCT and was conducted in the 2008-2009 school year in the same two Florida school districts with the same population of students and teachers as Phase 1. The specific questions addressed in Phase 2 were whether students whose teachers used *PCI* achieved (1) higher sight word reading scores and (2) higher phonological assessment scores than students whose teachers used their existing reading programs. PCI Education was also interested in whether the program's impact on sight word recognition was mediated by the amount of time teachers spend teaching those skills. Additionally, we investigated whether effects of *PCI* differed for specific subgroups of students: those who scored lower on sight word or phonological pretest, those in lower or higher grades, those who had teachers with more experience teaching Special Education, autistic students, and students who were English Language Learners. As an initial research base for the *PCI Reading Program*, this efficacy study was designed to determine whether students who are exposed to *PCI* learned more of the specific sight words taught in the program than students who were not exposed to the program.

Teachers who participated in Phase 1 were asked to use the program in Phase 2. We also tracked the students who had been exposed to *PCI* in Phase 1 and asked their Phase 2 teachers to use the program. We recruited additional teachers, who used their existing reading program and served as the comparison group. We were therefore able to extend the Phase 1 experimental design and analysis to use a matched quasi-experimental design as well as use an extra-experimental approach to estimate the impact of *PCI* after two years. The latter method used the first year gains of the *PCI* group to estimate the two-year impact given the former control group received *PCI* in Phase 2.

Findings. In both the quasi-experimental and extra-experimental approaches to estimating the two-year impact of *PCI*, we found that students in the *PCI* classrooms achieved significantly higher scores on the sight word assessment than students who were not exposed to the program. The difference found in the quasi-experiment (adjusted effect size of 0.89 with a .06 *p* value) was equivalent to a difference of 31 percentile points, and the difference found in the extra-experimental approach (adjusted effect size of 0.98 with a *p* value of .02) was equivalent to a difference of 34 percentile points. With a second year of exposure to the program, we found that students continue to improve their sight word recognition and that the effect of *PCI* was larger after two years than it was after one year. We did not report the impact of *PCI* on phonological skills because very few students progressed to *Level Two*- the program level in which phonological skills are introduced. Additionally, because we did not collect individual student usage data, we were not able to examine whether the impact of *PCI* on sight word recognition was mediated by the amount of time teachers spent teaching those skills.

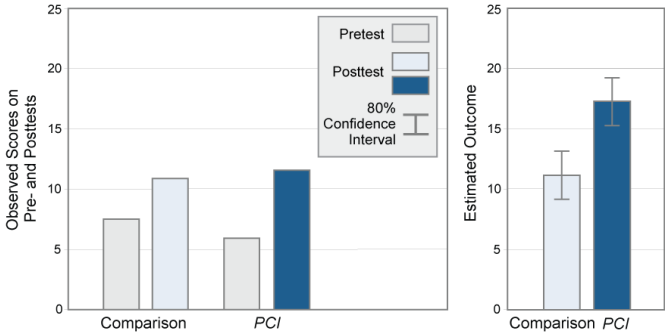

Figure 1. Impact on Sight Word Recognition Using Quasi-Experimental Approach: Unadjusted Pre- and Posttest Means for Comparison and *PCI* (Left); Adjusted Means for Comparison and *PCI* (Right)

When examining moderator effects using the quasi-experimental approach, we found that the sight word pre-assessment was not a significant moderator of the impact on sight word post-assessment scores. We have strong confidence that students whose teachers have more than four years of

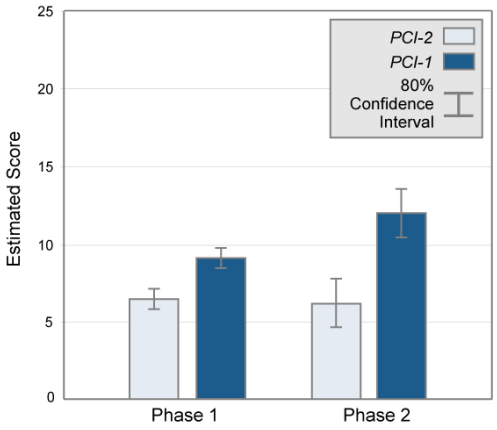

Figure 2. Impact on Sight Word Recognition Using Extra-Experimental Approach: Year 1 Impact (Left); Year 2 Impact (Right)

Special Education teaching experience benefit more from *PCI* than students with teachers who have fewer than four years of Special Education teaching experience. Due to the sample size and imbalance between the two groups on the phonological pre-assessment, these were the only moderators we were able to examine with this approach.

Using the extra-experimental approach, we were able to examine the moderator effects of the sight word and phonological pre-assessment and teachers' years of teaching Special Education. While we found no significant moderating effects, it is important to note that these analyses may be underpowered, given the small sample sizes in the program and control groups, and deserve additional exploration.

Overall Teacher Impressions. Qualitative data obtained from surveys, observations, and informal interviews show that, as in Phase 1, teachers were very satisfied and students were highly engaged with the program. Almost all of the teachers in the *PCI* group reported that they would continue to use the program after the study is over. Both teachers and administrators were encouraged that *PCI* fulfilled the need for a reading program specifically designed for this population of students. However, teachers reported that the primary difficulty in implementing the program was finding the time for the individualized instruction components of the program. Many teachers in the *PCI* group also reported using additional, supplemental reading materials. Moreover, student progress through the program was much slower than expected by the program developers—only half of the students learned more than 20 words.

Design and analysis. The study used a matched quasi-experimental design, comparing assessment scores of 26 students who had received exposure to *PCI* for two years to 51 students who had received no exposure to *PCI*. We also used an extra-experimental method to estimate the two-year impact of *PCI*, which compared scores of 28 *PCI* students who were part of the randomized *PCI* group in Phase 1 to scores of 12 Phase 1 control group students who used *PCI* in Phase 2. Multi-level analysis (hierarchical linear modeling) was used to estimate the program impact and the moderating effect of relevant variables. It takes into account the hierarchical nature of the data where student data were grouped within teachers. Information on student and teacher background characteristics as well as program implementation was gathered through online surveys, observations, and teacher interviews. The impact estimates were adjusted for any chance imbalances on relevant characteristics about students and teachers between the two groups. The two complimentary methods produced consistent impact estimates which provided us with convergent validity and greater confidence in our results.

Conclusion. This study provides evidence of the efficacy of the *PCI Reading Program*. The significantly large impact after two years found in both analytic approaches and high levels of teacher satisfaction with the program provides useful information for school districts looking for a reading program for severely disabled students. However, as we continue our research of the *PCI Reading Program* in both districts over the next three years, it will be equally important to examine why student progress is slower than expected.

Comparative Effectiveness of PCI Education's *PCI Reading Program*:

A Report of a Comparison Group Study in Brevard Public Schools and Miami-Dade County Public Schools

Table of Contents

Introduction

PCI Education has contracted with Empirical Education Inc. to conduct a five-year longitudinal study to determine the comparative effectiveness of the *PCI Reading Program (PCI)* as implemented in two school districts. The study aims to detect differences in achievement between students who have been exposed to *PCI* and students who have not been exposed to the program. We report here on Phase 2 of the research, which began in May 2008 in Brevard Public Schools (BPS) and Miami-Dade County Public Schools (M-DCPS) on the program's efficacy for reading achievement among students with severe disabilities. PCI Education initiated the research. The second year of the study (Phase 2) builds upon a randomized control trial conducted during the 2007-2008 school year in BPS and M-DCPS (Phase 1) which studied the efficacy of the *PCI Reading Program–Level One*.

The specific question we address throughout this five-year study is whether students whose teachers have been given *PCI* achieve higher sight word reading and phonological assessment scores than students whose teachers do not have the program. We are also interested in whether the impact of *PCI* on sight word recognition is mediated by the amount of time the teachers spend teaching those skills. Additionally, we are interested in whether *PCI* had a different effect for specific subgroups of students: those who score lower on sight word or phonological pretests, those in lower or higher grades, those who have teachers with more experience teaching Special Education, autistic students, and students who are English Language Learners. PCI Education also requested that we investigate the extent of program implementation and the level of interest generated among the teachers. Because this is an efficacy study, the program's effectiveness is not tested against standardized measures of reading. As an initial study of this program, our goal is to determine the extent to which *PCI* helps severely disabled students learn the specific skills on which the program focuses.

Description of Phase 1

Phase 1 of the study was a randomized experiment in which we investigated whether students whose teachers were given the *PCI Reading Program–Level One* achieved higher sight word assessment scores 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 Brevard Public Schools and Miami-Dade County Public Schools. Our outcome measure was a sight word assessment developed by an independent consultant. The experiment included 35 teachers (20 *PCI* and 15 comparison group teachers) and 128 students. We found that students in *PCI* classrooms performed significantly higher on the post-intervention sight word assessment than students in control classrooms.

For the results of this first phase, please see *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* (Toby, Ma, Jaciw, & Cabalo, 2008).

Description of Phase 2

Phase 2 of this study extends the analysis of outcomes for students involved in Phase 1 and also uses a comparison group. In all, 21 teachers agreed to be trained and to use the *PCI Reading Program* with their students (*PCI* group) and 13 teachers agreed to serve as the comparison group. Seventeen of

the 34 Phase 2 teachers had also participated in Phase 1 of the study.[1] We estimate the two-year effect of *PCI* in two different ways using appropriate subsamples of teachers and students.

We designed our study to provide useful information to support local decisions that take into account the specifics of district characteristics and their implementation of the PCI program. The results should not be considered to apply to school districts with practices and populations different from those in this experiment. This report presents a description of the conditions of program implementation and provides the reader with an understanding of the context of both district sites.

Phase 2 Research Questions

We identified the following questions in advance of the study. The questions pertain to the students and teachers to whom we can generalize our results:

1. What is the impact of the *PCI Reading Program* on student achievement in sight word recognition?
2. What is the impact of the *PCI Reading Program* on student achievement in phonological awareness (as measured by the DIBELS Initial Sound Fluency assessment)?
3. Is the impact of the *PCI Reading Program* on sight word recognition mediated by an intermediate impact of *PCI* on time spent teaching those skills?
4. Is there a differential effect of performance on sight word recognition based on students' scores on (a) the sight word pretest or (b) the phonological pretest?
5. Is there a differential effect of performance on sight word recognition based on students' grade level?
6. Is there a differential effect of student performance on sight word recognition based on their teacher's years of experience teaching Special Education?
7. Is there a differential effect of performance on sight word recognition based on students' disability classifications, specifically for students with autism?
8. Is there a differential effect of performance on sight word recognition based on students' English language learner status?

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

Methods

Our experiment is a comparison of outcomes for teachers who used *PCI* and for a comparison group of teachers using their district's current materials and methods. This section details the methods used to assess, with some level of confidence, the size of the average difference in performance between the groups. With this kind of study, we have to keep in mind that even where we detect a difference, factors other than *PCI* may contribute to the observed difference. We begin with a description and

[1] Thirty-eight teachers originally consented to participate in Phase 2 (22 *PCI* and 16 comparison teachers). However, four teachers left the study after consenting to participate (one *PCI* and three comparison teachers). The reasons for this attrition are further described in the Teacher Attrition section of this report.

rationale for the experimental design and then describe the intervention, the research sites, the sources of data, the composition of the experimental groups, and the statistical methods used to generate our conclusions about the impact of *PCI*.

Experimental Design

Experiments are used to estimate the impact of an intervention on the basis of a sample drawn from a larger population. Members of the sample are selected in such a way that the impact of the intervention—in this case, on the selected teachers and students in the sample—would also be expected to apply to the larger population from which the sample was drawn. Here we used a convenience sample of teachers and students rather than a formal random sample.

The design of the experiment attempts to reduce bias and imprecision and to make our impact estimates (based on the sample) as accurate and precise as possible. Still, there is always a level of uncertainty. We think of the uncertainty as related to the likelihood that we would get a different result if we took a new sample of students or teachers from the same hypothetical population. Uncertainty also increases when treatment is confounded with other variables that affect performance, and this confounding is not controlled for. Our design attempts to efficiently deploy the available resources to reduce uncertainty and improve both accuracy and precision.

In this study we use two experimental approaches to obtaining estimates of the impact of the program. The first is a quasi-experimental approach that involves matching students who receive *PCI* to a comparison group who has never been exposed to the program. The discussion in this section is concerned with the process of identifying the program and comparison samples as well as specification of the statistical equation that is used to estimate the program effect. The second is an 'extra-experimental' approach that is attributable to Bell and Bradley (2008). With this method we work with the original sample from Phase 1 of the experiment, in particular, the subset of students who were parts of the originally randomized groups and who continued in the study through the end of Phase 2.[2] The approach considers the difference between the experimental groups (those originally assigned to either *PCI* or the control condition) after one year of the experiment, as well as the difference between them one year after that initial year (i.e., after the original program group has received two years of exposure to *PCI*, while the original control group has received one year of exposure following a year without exposure). Importantly, the estimate of the two-year impact that is obtained using these two difference estimates is based on a solid experimental foundation: the estimates that are components of the extra-experimental estimate are themselves unbiased estimates of impacts, which means that the estimates cannot be biased by selection or secular trends affecting both experimental groups. The method does assume, however, that the intervention itself does not change from one year to the next. The *PCI Reading Program* did not change from Phase 1 to Phase 2, which makes the extra-experimental estimate appropriate for this study.

As part of our research process we specify the design or plan for the experiment at the outset. This pays off in two ways. First, we identify, before seeing the outcomes, where we expect to detect an impact and which factors we suppose will moderate that impact. Thus we specify the research questions in advance. In this way, we avoid fishing for results in the data, a process that can lead to mistaking chance differences in outcomes for differences that are probably important as a basis for decisions. In other words, because some differences will appear large simply by chance, "mining" the data can lead us to conclude that an effect exists when it does not. While we can still conduct exploratory analyses after establishing main effects, these are useful mostly for generating ideas

[2] With the Bell and Bradley approach we are considering single- and multi-year exposures of students, not teachers. Many of the teachers who were assigned to the control condition at the start of the study received *PCI* in the second year, with a few exceptions (e.g., one teacher insisted on remaining in the control condition.)

about how the new program worked—that is, as hypothesis-generating efforts for motivating future study—rather than as efforts from which we draw firm conclusions from our existing study.

Second, an experimental design will include an analysis of how large the study should be in terms of students, teachers, and schools in order to obtain the desired level of confidence in the results. In the planning stage of the experiment, we calculate either how many cases we need to detect a specifically sized difference between the *PCI* and comparison groups, or how big a difference we can detect given the sample size that is available. Technically, this is called a power analysis. We will explain how many aspects of design determine the size of the experiment.

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. One level in the hierarchy is identified as the level or unit of analysis and 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 comparison and *PCI* units. In this study, students will represent the primary unit of analysis and our generalization will extend beyond these particular teachers to others who could, in principle, replace them within these schools. Appropriate analytic models are used to control for clustering of students in teachers. The outcome measures are student-level test scores on the sight word post-assessment that was developed specifically for this study and the phonological post-assessment, which was a modified version of DIBELS Initial Sound Fluency assessment.

What Factors May Moderate the Impact of *PCI* on Sight Word Recognition?

While the outcome measures include both the sight word and phonological assessments, the program levels being used in Phase 2 focus primarily on sight word skills. Therefore, we focus only on factors that may moderate the impact of *PCI* on sight word recognition.

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, students whose teachers have more than four years of Special Education teaching experience, students with autism, and students who are English Language Learners. 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 a particular moderator changes.

What Factors May Mediate Between *PCI* and Sight Word Recognition?

We have also identified variables that were to be measured after the experiment started and are believed to facilitate the student outcomes. In this experiment, we will measure the average reported number of minutes of sight word reading instruction from both the *PCI* and comparison teachers. This variable is itself a measurable outcome of the *PCI Reading Program*. Technically, these variables are called mediators. A mediator lies along the causal path between the point where we assign cases to the *PCI* or comparison group and the point when we measure student performance after the intervention is over. A mediator can either block or enhance the effect of an intervention, entirely or in part.

We usually think of a mediator as part of how the program has its impact. Our experiment can determine whether the program caused a difference in the mediator and the final outcome. We can

then use this information to draw a further conclusion about whether the difference in the final outcome is facilitated through a change in the mediator.

A limitation of mediation analysis is that we do not assign cases to levels of mediation, and therefore we cannot be sure that it is the factor that we identify as a mediator that truly facilitates the effect; in fact, it could be some other factor that is correlated with the variable that we identify that is the true mediator. The mediating variable and the final outcome do not have to be at the same level. In this study, the mediating variable is measured at the teacher level and the final outcome, student achievement, is measured at the student level.

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 a specific level of confidence that the intervention has an impact of a certain size. Or, as in this experiment, to decide whether the number of available students and teachers is adequate to detect the impact. This is an important part of experimental design and here we describe the factors we 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 higher. 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 we decided to set the MDES at 14 percentile points or, in terms of the standard deviation units we introduce below, an effect size of 0.35. (This was our starting value, but the MDES we can detect is determined by the numbers of teachers and students that are actually available at the point of analysis.)

How Much Variation Exists Between Teachers?

We are primarily concerned with differences in performance between students in program (*PCI)* and comparison classes. We focus on the uncertainty resulting from a hypothetical resampling of teachers and students for the schools in the study. In this instance we pay special attention to the differences among teachers in average student outcomes. Since the number of teachers is smaller than the number of students, much of the uncertainty in our estimates will be driven by the variation among teachers. The greater the variation among teachers in student-average scores, the more units 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 are large and/or the differences within the group of 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 small so that a very large proportion of the variation is attributable to differences among students they teach, then the sample size that matters most is the number of students. A summary statistic, 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 teacher-averages of the student outcome to the total variation in outcome. We assume that this is computed before the intervention. For this experiment we assumed a fairly conservative intraclass correlation of 0.20.

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, 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 (0.80).[3]. 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. In a study where we use a matched comparison group, including the pretest in the analysis also has the potential to reduce bias. Often there is imbalance between conditions on the pretest, and including it as a covariate eliminates bias resulting from this imbalance. (Conditioning on a pre-intervention measure of performance is shown to be especially effective at reducing bias in the effect estimate.)

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 unit of assignment—that is, where each unit has some portion of that subgroup—the impact on the power analysis is minimal. However, if our subgroup of interest is a subtype of the unit of assignment, then we expect to have less statistical power to measure the differential effect. In the current experiment, we are interested in differences between students at different levels of prior achievement, students at different grade levels, students with autism as compared to students with intellectual or other disability classifications, students who are designated as English Language Learners (ELL) as compared to non-ELLs, and differences in impact for teachers with fewer than four years of teaching experience in Special Education as compared to teachers with more experience. Only the last of these requires us to divide the sample of teachers to determine whether the impact varies among subtypes of teachers; thus the reduction in power described above applies to this analysis.

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

We express the uncertainty inherent in our result by figuring in the variation in the outcome that we would expect if we ran the experiment again with a hypothetical different sample of students and teachers from the same district. Although results are never exactly identical, we can design the experiment so that the various results we would obtain 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

[3] That is, we assume that .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.

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 obtaining a result like that 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 46 teachers (23 in each experimental group) would constitute a sufficiently large sample to detect an impact of 14 percentile points (or an effect size of 0.35) in Phase 2. Thirty-four teachers (21 *PCI* teachers and 13 comparison teachers) were recruited for this phase of the experiment. Only 24 of these teachers were included in the analytic sample (of the quasi-experimental analysis: 11 *PCI* teachers and 13 comparison teachers). Given the obstacles in recruiting a sufficient number of teachers, we anticipated being able to detect an effect size of 0.51 rather than the 0.35 originally planned for. The experiment is underpowered with respect to the original goals; however, given the overall reading effect size found in Phase 1 (0.55), we felt that the experiment is still within a reasonable design size.[4] Data collected from subsequent phases of the study can be added to the current sample in order to achieve greater statistical power if needed.

Method of Analysis

We used two strategies to limit the effects of selection bias in the estimation of the two-year quasi-experimental impact estimates. The first is a matching strategy, whereby for each program case we identified comparison cases who were comparable in terms of specific background characteristics. The second strategy involves statistically adjusting the impact estimate to control for imbalance between conditions on the same characteristics. We describe each approach below:

Matching Process

The goal was to find comparison cases similar to the program cases on background characteristics that are likely to affect performance and that, if imbalanced between conditions, could bias the impact estimate.

Ideally matching should be performed in terms of characteristics that were measured before the start of the study (technically, before cases are assigned to conditions by the given selection mechanism). The recruiting schedule prevented us from collecting background data for the outside comparison group prior to Phase 1. This means we collected background information for the two-year experimental sample at the start of Phase 1, and for the comparison group at the start of Phase 2. We do not regard this as too problematic because the information for the comparison group used for matching is unlikely to be affected by earlier exposure of other cases to the program *(PCI)*. In other words, the characteristics used to form the matches are highly unlikely to be influenced by the program.

We used a form of matching called propensity score matching. This involves modeling the probability of a student being in the program or comparison group using a set of variables that

[4] In a later section of the report, we show that the effect of *PCI* on sight word performance is large and is detectable despite the small teacher sample. Additionally, because we focused on the two-year impact of *PCI,* data from some teachers and students were not used in any analysis. It was important to collect data from these students for potential use in for future years of the study.

potentially influence selection into the two groups. Each student in the program condition is matched to one or more students who have a similar propensity score. In our study, our comparison group was relatively small; therefore, we had to figure in the trade-off between excluding comparison cases that are not well matched and loss of statistical power. Our approach was to exclude comparison cases whose propensity scores lay outside the range of the propensity scores for the treatment cases. This led to the removal of several comparison cases for whom matches had never been made. We divided the propensity scores for the treatment group into quartiles and then, within each quartile, checked for balance between the treatment and comparison cases simultaneously on the covariates used to calculate the propensity scores.[5] Balance was achieved within each quartile (though as we note below, this favorable result may have been due in part to the relatively small number of cases in each quartile). We then used all of the program cases and the comparison cases who were not eliminated from the pool through the matching process for further statistical analysis.

Statistical Adjustment

With propensity score matching we eliminated a small number of comparison cases whose propensity scores were outside the range of the treatment group's propensity scores. We demonstrated balance between conditions within quartiles of the propensity score on several covariates. Given the sample sizes, we were concerned that the statistical power for detecting differences between conditions in the balance checks was low, and could therefore indicate balance in a situation where, with a larger sample, we would find imbalance. To further adjust for possible imbalance between conditions on specific covariates and thereby reduce the likelihood of selection bias influencing results, we conditioned our impact estimate on these covariates. That is, we adjusted our effect estimates to account for possible imbalance on these covariates by including them in the statistical equation. We also included the propensity score from the matching stage as a covariate. This strategy is recommended by Shadish et al. (2006).

Intervention

The intervention we are testing in Phase 2 of this study consists of the *PCI Reading Program–Level One* and *Level Two* kits and a one-day training for the teachers.

Training/Professional Development

PCI teachers from both sites were invited to participate in a one-day training to become familiarized with the *PCI Reading Program–Level One* and *PCI Reading Program–Level Two*, the pre-assessment administration, and the research study. Teachers who used *Level One* during Phase 1 and teachers who had never used *PCI* attended the same training. Comparison teachers in BPS were also invited to participate in a half-hour long training to become familiarized with the pre-assessment administration and research study. Trainings for the two districts occurred at separate times and locations.

In M-DCPS, 8 out of 11 *PCI* teachers attended the training. Five of the eight teachers had been participants in Phase 1 (two in the *PCI* group and three in the control group) and three were new to the study. Also in attendance were three district support staff members and the point-of-contact

[5] We used quartiles instead of quintiles because, for one of the quintiles, the number of comparison cases with propensity scores that lay within that quintile was very small and would not have allowed a balance check.

(POC) for a portion of the day. Of the three teachers who missed the training, one had been trained in Phase 1.The two remaining teachers received the opportunity for a follow-up training with a district support staff member but were not trained.

In BPS, all 10 *PCI* teachers attended the training. Nine of these teachers were participants in Phase 1 (five in the *PCI* group and four in the control group) and one teacher was new to the study. Also in attendance was a teacher of a new autism unit whose principal specifically requested that she be trained in and use *PCI* with her students. This teacher will be excluded in the study to avoid introducing outside bias. The POC in BPS also attended the training.

Those who attended the training were trained in *Level One* in the morning and *Level Two* in the early afternoon. They then stayed for a half-hour session at the end of the day to discuss the study and its expectations, as well as to be trained in how to administer the pre-assessment. The comparison teachers from M-DCPS did not attend the training at the end of the day, due to limits on the availability of substitutes, according to the POC. District staff who had attended the training later trained the comparison teachers in how to administer the pre-assessment. In BPS, all six comparison teachers were trained in the pre-assessment and discussed the study as they arrived (either individually or in pairs). In both districts, the majority of the half-hour training was led by a researcher from Empirical Education.

The *Level One* and *Level Two* training, provided 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 previous research as well as to the literature and rationale behind the *PCI Reading Program*. In both districts, the training began with a theoretical and pedagogical overview of *PCI*. Haney explained the student prerequisites needed prior to starting each level and for moving forward within each level. Haney emphasized that the program should be implemented with a high level of fidelity to the Teacher's Guide. For the remainder of the training, Haney modeled the lesson cycle, explained the new elements of *Level Two*, discussed ways to adapt the program for non-verbal students, and described different implementation techniques so that teachers would be able to work within the context of their own classrooms. Feedback from teachers who had used the program during Phase 1 was also incorporated into the training.

In M-DCPS, *Level One* and *Level Two* kits were distributed to teachers who needed them. Teachers were able to practice the lesson cycle steps with the materials. In BPS, the kits did not arrive in time for the training. Haney used the POC's *Level One* kit, but did not have a *Level Two* kit to model with. The kits were mailed to the BPS teachers within one week of the training.

PCI Reading Program Materials

The *PCI Reading Program* 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 foundation of the program is its bridging approach of the three levels to teaching non-readers how to read. Non-readers begin with *Level One*, which aims to teach students 140 sight words and common nouns and verbs through visual discrimination. *Level Two* aims to teach 140 additional words as well as a few commonly used inflection endings, such as –s and –ing. In *Level Three* (which is still under development and not included in Phase 2 of the study), the 280 sight words are linked together by phonetic patterns to develop students' basic decoding strategies and word-attack skills. 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. This is followed by 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 for each level. The program includes reproducible sheets for parents to work on with their children.

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. The *PCI* trainer agreed that the use of supplemental materials was allowed as long as teachers used *PCI* for the recommended time. Teachers are expected to follow the curriculum and lesson cycles, 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. However, the *PCI* trainer acknowledged that the minimum number of minutes may be a more realistic implementation expectation.

District Materials

In this study we compare *PCI* instruction to the customary teaching approach without *PCI, or* "business as usual." To gain an understanding of the existing reading materials in each district, researchers included questions on the teacher background form in the participant information packet regarding the materials teachers had used during their prior school year. All of the teachers in both districts who had used *PCI* during Phase 1 and were continuing their participation during Phase 2 (a total of eight teachers) indicated on this questionnaire that they had used *PCI*. Two of these teachers also indicated having used a supplemental reading program.

Among the 11 teachers in BPS who had not used *PCI* during Phase 1, 64% reported using EdMark with their classes in addition to, or supplemented by, other programs. No teachers reported using EdMark as the sole method of instruction. Other programs/instruction methods used by two or more teachers include News-2-You, Dolch/Fry sight words, SRA Corrective Reading/Reading Mastery, Scott Foresman Reading for Florida (depending on student level), and McMillan/McGraw-Treasures, and Triumphs.

Of the 19 teachers in M-DCPS who had not used *PCI* in Phase 1, 42% reported using Houghton Mifflin with their classes, and of these teachers, three-fourths report using Houghton Mifflin as their sole method of instruction. Two teachers in M-DCPS reported using EdMark and three teachers reported using The Letter People. Overall, most teachers used a variety of sources for instruction.

Schedule of Major Milestones

Table 1 lists the major project milestones and associated dates. Planning for Phase 2 began in May 2008. The remaining milestones will be described in the Participant Recruitment and Data Collection sections below.

Table 1. Research Milestones: Phase 2

Date	Milestone
May 2008	Initiation of the Phase 2 project
May–October 2008	Recruitment of school districts, teachers, and assessment consultant
July–October 2008	Development of assessment and district approval
October–November 2008	Question-and-answer sessions and training
October–November 2008	Administration of pre-assessments, start of implementation, and initiation of monthly web surveys
March 2009	Classroom observations
May–June 2009	Administration of post-intervention assessments and completion of data collection

Participant Recruitment

District Identification

Empirical Education researchers contacted the person designated as *PCI*'s POC in each district during Phase 1 of the study and explained the details and procedures of continuing the study. Both agreed to continue to act as the POC during Phases 2 through 5 and identified eligible teachers who met the criteria to participate. However, due to hurricanes/weather conditions and other administrative issues within the districts, teacher recruitment was delayed. Empirical Education secured notification of acceptance of our research application from M-DCPS on July 30, 2008 and a signed district agreement from BPS on October 20, 2008.

Teacher Identification

In Phase 2, the teachers were not randomly assigned to the *PCI* group or the comparison group. Teachers who taught reading to students with supported level disabilities in grades 3-8 (target students) from the Miami-Dade and Brevard districts were identified by district staff and invited to participate in the study. The following criteria determined if a teacher would be eligible to teach *PCI* in Phase 2:

- If the teacher participated in Phase 1 of the study: As an incentive for participation in Phase 1, all teachers were promised *PCI* training and materials the following school year. Therefore, Phase 1 teachers who taught target students in Phase 2 were eligible to teach *PCI* in Phase 2.
- If the teacher taught students who received *PCI* instruction in Phase 1: As part of the study design, researchers will attempt to follow students as they progress through the program. Therefore, teachers in Phase 2 who taught students who received *PCI* instruction in Phase 1 were eligible to teach *PCI* in Phase 2.

All other teachers of target students were invited to participate in the study as part of the comparison group. Comparison teachers will continue to use their current district materials until the 2010-2011 school year, when they will also be given the *PCI Reading Program*. The validity of the

inference concerning the effectiveness of the program depends on the comparison group being similar in all important respects to the group that receives *PCI*. Where there are differences, we can to some extent control for them statistically. On a priori grounds there is no reason to believe that the *PCI* and comparison groups are so fundamentally different that they cannot be compared without substantial bias affecting the results.

Based on our research goals and other estimates of *PCI*'s likely impact, we anticipated that we would need a sample size of approximately 46 teachers (23 in the *PCI* group and 23 in the comparison group). In this case, the sample was identified based on the number of teachers who met a set of criteria. The first criterion was that teacher must teach students with supported level disabilities, since the program is designed for students with intellectual disabilities and autism. The second was that teachers must teach students who are in 3rd through 8th grade. At the request of the publisher during Phase 1, researchers attempted to limit the 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. During Phase 2, researchers decided not to extend the sample to include high school grades due to the differences in how reading instruction is conducted in high schools. Finally, teachers must teach a self-contained reading block. 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.

Teacher Recruitment

Researchers provided district staff members with the names of teachers and students involved in Phase 1 of the study. The district provided researchers with each student's school and teacher during the 2008-09 school year (Phase 2). Researchers and district staff were then able to identify which teachers would be invited to participate in Phase 2 as *PCI* and comparison teachers.

Because the next phase of recruiting happened internally with the district POCs contacting all eligible comparison teachers, researchers do not have information on the full population of teachers and are unable to determine how many teachers were invited to participate. We do know, however, that one teacher in M-DCPS met the criteria to teach *PCI* because she had been a control teacher in Phase 1, but requested to continue to use her existing reading materials and participate in Phase 2 as a comparison teacher. We included this teacher in the comparison group because neither she nor her students had exposure to *PCI*.

All identified teachers were sent an informational flier briefly describing the study and a participant information packet which included four elements:

- Description of participant responsibilities
- Study timeline/overview
- Research participant agreement form
- Teacher background/contact information questionnaire, which was filled out and returned to researchers

Researchers also offered to host a voluntary telephone question-and-answer session for all interested teachers. Teachers were notified about these sessions through the informational flier. These sessions provided a format for us to describe the specifics of participation in the study as well as to answer potential participants' questions and to address their concerns. Each district was offered two different dates/times to call in. No teachers from M-DCPS called in to either session. One teacher from BPS called in to the first session and approximately seven BPS teachers called in to the second session.

All teachers who met the criteria to use *PCI* (those who were in Phase 1 of the study or those who had students who received *PCI* instruction during Phase 1 of the study who taught students with

supported level disabilities in grades 3-8) were invited to attend a one-day training on *PCI Reading Program–Level One* and *PCI Reading Program–Level Two* on October 16, 2008 (M-DCPS) and on November 11, 2008 (BPS).

Student Identification

Within the study classrooms, not all students are appropriate candidates for the *PCI* program. In addition to the recommendation that *PCI* be implemented for non-readers with developmental disabilities, autism, and significant learning disabilities, *PCI* also designates that, prior to using the program, students must be able to:

- Follow simple, one-sentence directions
- Demonstrate their understanding of a teacher request by either pointing or responding verbally
- See words on a page and somehow point to or otherwise indicate identification of those words
- Communicate a response to a question or directive
- Visually discriminate between words and letters (they do not need to know the alphabet)

Therefore, researchers asked the *PCI* and comparison teachers to adhere to these prerequisites.

Site Descriptions

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, 2006).

BPS has 130 schools serving pre-kindergarten through grade 12. The total enrollment is 74,807 students (Florida Department of Education, 2008). Table 2 and Table 3 provide information about the entire district.

Table 2. Demographics of Brevard Public Schools

Brevard Public Schools	
Total schools	130
Total teachers	4,888
Grade structure	PK-12
Student enrollment	74,807
Percent of students designated as:	
Disabled	17.1%
English language learner	2.1%
White	71.0%
Black	14.3%
Hispanic	7.7%
Asian/Pacific Islander	1.9%
American Indian/Native Alaskan	0.3%
Multi racial	4.8%
Source: Florida Department of Education, 2008	

Miami-Dade County Public Schools

Miami-Dade County Public Schools (M-DCPS) 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, 2006).

M-DCPS has 441 schools serving pre-kindergarten through grade 12. The district's total enrollment is 353,831 students (Florida Department of Education, 2008). Table 3 provides information about the entire district.

Table 3. Demographics of Miami-Dade County Public Schools

Miami-Dade County Public Schools	
Total schools	445
Total teachers	22,055
Grade structure	PK-12
Student enrollment	353,831
Percent of students designated as:	
Disabled	11.6%
English language learner	16.2%
White	9.4%
Black	26.9%
Hispanic	61.2%
Asian/Pacific Islander	1.2%
American Indian/Native Alaskan	<0.1%
Multi racial	1.3%
Source: Florida Department of Education, 2008	

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 and phonological assessment scores, student demographic data, and data from training observations, classroom observations, teacher surveys, informal interviews, email exchanges, and telephone conversations. In addition, we have reviewed various program documents and materials. Therefore the research employs a multiple methods approach through which we will 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 and unique identifier
- Date of birth
- School the student attends
- Classroom teacher
- Gender
- National School Lunch Program status (proxy for socio-economic level)

- Ethnicity
- English learner status
- Grade level
- Disability/eligibility codes

The students' school and teacher data were used to link all students to their teachers. Date of birth was used as an additional student identifier. All other data were used in the matching procedures or to conduct moderator analyses.

Researchers did not look at the differences in outcomes for students based on their disability classification in Phase 1. However, at the request of our *PCI* POC, Phase 2 will explore the differences between students with autism to those with other classifications.

District POCs told researchers that the state of Florida was moving toward collapsing the Educable, Trainable and Profoundly Mentally Handicapped codes (EMH/TMH/PMH) into an "Intellectually Disabled" code. This change is happening this year and will be changed in district databases as each student's Individualized Education Plan (IEP) is reviewed. This means that some students have the new code (*W* for Intellectually Disabled), while other students still have the old code (*A* for EMH; *B* for TMH; *N* for PMH)). While researchers were expecting these codes, we also received specific learning disabled, other health impaired, emotional behavior disability, orthopedically impaired, and dual sensory impaired codes. The *PCI* POC explained that while *PCI* is meant for students with intellectual disabilities or autism, the program can be appropriate for students with these other codes if their teacher feels it is appropriate.

M-DCPS requires that parents sign a consent form prior to release of data; therefore, researchers received data only for students who had returned signed parental consent forms. 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). Since researchers do not have access to full classroom rosters, we are unable to determine the percent of eligible students whose data are included in this study.

Empirical Education received these data from M-DCPS and BPS in January 2009.

Achievement Measures

Sight Word Pre- and Post-Intervention Assessment

The primary outcome measure is student assessment scores on the sight word post-assessment, which was initially developed by an independent consultant for Phase 1 of the study. Prior to Phase 1, the development specialist took the following steps to determine the appropriate words for both the pre- and posttests:

1. Selected only words that are taught in both the *PCI* and Edmark reading programs. (Prior to Phase 1, it was thought that Edmark would be the primary curriculum used in the control group; however, researchers learned that, in practice, teachers were using a variety of curricula.)
2. Used the EDL Reading Core Vocabulary Cumulative list to determine the reading levels of each word. This was important so each of the two 10-word tests 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 long term memory issues.

Prior to Phase 2, we sent the Phase 1 sight word list to an Associate Professor from Florida State University’s College of Education and the Center for Reading Research—an expert in the Special Education reading domain—who provided researchers with a list of words appropriate for the pre- and post-assessments for Phase 2. The newly recommended list included several words from the Phase 1 list to act as anchor items, as well as additional words from *Level One* and *Level Two* that were likely to appear in comparison reading programs. Empirical Education computed test statistics for the sight word assessments for both the pretest and the posttest in Phase 1. Both exhibited homogeneity and high internal consistency, as well as correlations between the different versions of the sight word assessments (pre- and post-). The anchor items used in the previous version of the assessment were used in the post-assessment as well. Additional words were randomly selected from the *Level One* words provided by the expert from Florida State University. In the end, however, we decided to not use the *Level Two* words on the post-assessment because of the small number of students (six) who had used *Level Two* during this school year.

Phonological Pre- and Post-Intervention Assessment

In addition to the sight word assessment, the phonological post-assessment serves as an outcome measure. The phonological pre-assessment was also developed by an independent consultant during Phase 1. The same assessment items used as a pre-assessment in Phase 1 were used as a pre-assessment in Phase 2. The phonological pre-assessment, administered to all participating students at the beginning of Phase 2, 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 pre-assessment, teachers are 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.

During Phase 1, teachers and students had used only *Level One*, which does not include phonics instruction. Therefore, we had not included phonics achievement as an outcome measure. During Phase 2, however, the phonological assessment will be used as an outcome measure so that researchers will be able to track improvement in phonics as students begin to receive phonics instruction with *Level Two* and future levels over the next four years. The Associate Professor from Florida State University who provided words from the sight word pre- and post-assessment for Phase 2 was again contracted to advise researchers on the selection of an appropriate phonological post-assessment. Recommendations included:

- Woodcock Johnson III (Letter-Word Identification, Word Attack, Picture Vocabulary subtests)
- Comprehensive Test of Phonological Awareness (CTOPP) developed by R. Torgesen and Wagner (Sound Matching and Blending subtests)
- AIMSweb Letter Sound Fluency
- Phonological Awareness Literacy Screening (PALS)
- DIBELS Letter Naming Fluency
- DIBELS Initial Sound Fluency

After reviewing the cost, testing time, and appropriateness of each assessment, researchers selected the DIBELS Initial Sound Fluency as the phonological post-assessment measure. Researchers contacted Dynamic Measurement Group (DIBELS publishing company) and requested permission to use the assessment as part of the research and to modify the standard administration and scoring procedures.[6] This request was approved for our research with this population of students.

Pre-Assessment Training and Post-Assessment Procedures

Teachers involved in Phase 1 of the study had administered the sight word and phonological pre-assessments in the previous year. While the other teachers had some experience in administering sight word assessments, they were not already familiar with the specific administration of the assessments described above. In the pre-assessment training, expectations of the program and the study were reiterated to both *PCI* and comparison teachers. Trainers emphasized to *PCI* teachers that the pre-assessments must be administered before instruction in the program begins.

As described in the Experimental Design section, teachers from both districts received a list of student prerequisites to determine which students should be assessed and included in the study. Teachers from M-DCPS were emailed the list following the training, and teachers from BPS received the list in their pre-assessment packet/training. Researchers followed up with survey questions which asked how teachers determined which students to assess.

For both districts, the pre-assessments were administered between October 2008 and January 2009, and all post-assessment were conducted between May and June 2009. Teachers were instructed to mail completed pre-assessments to Empirical Education in postage-paid envelopes.

The post-assessment packets were mailed to all teachers in early May 2009. While a formal training on how to administer the phonological post-assessment was not provided, teachers received detailed instructions in the post-assessment packets and were encouraged to contact researchers if they had any questions. Teachers received the following materials in their post-assessment packet:

- A cover letter describing the packet contents and directions for returning the assessment items (one per packet)
- A teacher questionnaire (which was completed for each individual study student) asking if the pre- and post–assessments were administered and if the student was classified as non-verbal at the beginning and end of the study. *PCI* teachers were also asked which level and word students were on at the beginning and end of the year, which will be used to describe implementation (one per student).
- Sight word assessment administration, scoring guide and word card set (one per packet)

[6] Modifications were required because the assessment is not expressly designed for Special Education students. These modifications consisted of the following: instructing the teachers to administer the assessment untimed (which was recommended by the FSU expert) and to use the raw score (total correct) rather than calculating the score based on the response time of the student (which is the standard scoring procedure).

- Sight word assessment scoring sheet (one per student)
- Phonological administration, scoring guide, and picture card set (one per packet)
- Phonological assessment/Initial Sound Fluency: short form directions/Progress Monitoring Sheet (1 per student)
- Additional Parental Consent Forms, if needed.
- Pre-stamped envelope to return materials (1 per packet)

Methods Used to Investigate the Intervention Implementation

In addition to quantitative data, researchers also collected qualitative data over the entire period of Phase 2 of the experiment, beginning with the teacher recruitment phase and ending with the academic calendar of the district in June 2009. Training observations, classroom observations, multiple teacher surveys, informal interviews, email exchanges, and telephone conversations are used to provide both descriptive and quantitative evidence of the implementation. Informal interviews, email exchanges, and telephone conversation were not scheduled and occurred as needed; however, information received during these exchanges was used to further inform observation and survey data.

Survey Schedule

Surveys were deployed to both *PCI* and comparison group teachers beginning in November 2008 and continuing on a monthly basis until May 2009.[7] Table 4 describes the survey schedule and response rates.

Table 4. Survey Schedule

Survey	Deployment	Response rate
Survey 1	November 7, 2008 (M-DCPS) November 14, 2008 (BPS)	100%
Survey 2	December 12, 2008	100%
Survey 3	January 9, 2009	100%
Survey 4	February 6, 2009	100%
Survey 5	March 6, 2009	100%
Survey 6	March 27, 2009	97%
Survey 7	April 17, 2009	97%
Survey 8	May 15, 2009	94%

Classroom Observations

In general, observational data were used to inform the description of the learning environment, study participation, instructional strategies employed by the teachers, use of curricular materials, and student engagement. These data were minimally coded.

[7] Due to the late start of the project and delays during recruitment, two surveys were deployed in March.

Classroom observations took place in both districts in March 2009. 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 to visit one-third of participating teachers and classrooms and to meet with a small sample of school principals. Class selection for observations was based on "stratified convenience."

In each district, we first selected schools containing multiple teacher participants in order to maximize the number of teachers observed within the allotted time. Next, we looked at whether we had a fair representation in our sample of middle and elementary classes, *PCI* and comparison classes, and teachers who have used *PCI* for two years and those in their first year of implementation. 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.

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

At M-DCPS, we visited 6 out of 15 schools and 8 out of 18 classes. At BPS, we visited 4 out of 12 schools and 7 out of 16 classes. This selection allowed us to visit 10 out of 27 schools and 15 out of the 34 classes in the study. Of the 15 classes, six were middle school classrooms, six were *PCI* classroom, and three teachers had used *PCI* for two years.

During these classroom visits, we observed how teachers designed and carried out instruction. Across both assignment groups, we were also interested in how teachers organized instruction—group work, individual work, one-on-one instruction—how other adults in the classroom interacted with students, and the level of student engagement. Specifically, in *PCI* classrooms, we documented teachers' use of *PCI* and other materials as well as how closely they followed the prescribed *PCI* curriculum. For the comparison group, we hoped to obtain an idea of what types of curricula were enacted across the classrooms and to understand the degree to which instruction was individualized for students. All classroom observations were conducted within a period of one week across both districts. One researcher from Empirical Education conducted the observations. The POC from PCI Education was present at most of the observations in M-DCPS but was unable to attend the BPS visits.

Once the classroom observation schedule was set, researchers contacted principals at four of the scheduled schools and requested a brief interview with them following the observation. These principals were selected based on time convenience (i.e., if the observer had time in between or after observations). The purpose of this interview was to provide researchers with information, from the principal's perspective, about supported level disability programs, available support, and, if used in their school, how *PCI* was working. Two principals responded that they were available and willing to be interviewed. One Empirical Education researcher conducted the interviews after observing the teachers at the school. While the information gathered during the interviews provided researchers with context about the schools and classrooms, data will not be presented in the results section due to the small number of interviewees.

Implementation Data, Collection, and Analyses

Table 5 lists the implementation areas of investigation for this study, the types of analyses that were conducted for each area, and the data sources. These components and the rationales for the analyses are explained below.

Table 5. Implementation Data and Analyses

Area of investigation	Types of analyses	Data source(s)
Teacher background	Balance checks, moderator analysis, and descriptive	Teacher surveys
Conditions for implementation	Compare the conditions under which the *PCI* and comparison programs are implemented	Observations, Teacher surveys, Email exchanges, Informal interviews, and Telephone conversations
Description of implementation	Compare implementation of *PCI* to that of the comparison programs measure the extent to which teachers meet *PCI*'s recommendations for implementation	Observations, Teacher surveys, Teacher questionnaire on student progress, Email exchanges, Informal interviews, and Telephone conversations
Correlations between levels of implementation and student outcomes	Measure the extent to which student achievement is correlated with the different levels of *PCI* implementation	Teacher surveys

Teacher Background and Classroom Description

This study collects teacher background data to provide a context for reading intervention implementation. Because recent literature correlates teaching experience and content knowledge with teacher quality (Amrein-Beardsley, 2006; The Center for Public Education, 2005), we conducted balance checks on teacher background data to establish comparability between the *PCI* and comparison groups. We also conducted a moderator analysis on teachers' years of experience. For descriptive purposes, we present additional characteristics about teacher background and descriptions of the classrooms.

The data reported include:

- Years of teaching experience and subjects taught
- Information about credentials and certification
- Education level completed

According to administrators in the study districts, supported level classrooms are categorized or classified differently depending on the students' abilities and disabilities. Researchers asked teachers in both groups to select the best description of their classroom. The descriptors include:

- Intellectual Disabilities at the Supported Level
- EMH: Educable Mentally Handicapped
- TMH: Trainable Mentally Handicapped

- PMH: Profoundly Mentally Handicapped
- ASD: Autistic Spectrum Disorder
- Varying Exceptionalities
- Other

Conditions for Implementation

To understand the implementation data, it is critical to have information about the context in which the implementation takes place. We compared the classroom implementation conditions between *PCI* and comparison groups, and the data are reported using descriptive statistics. Areas of investigation include program training, availability of program materials, and availability of teaching assistants.

Program Training

PCI Education is particularly interested in learning about the teachers' assessment of how effective the training was in preparing them to implement different components of the program. The survey posed identical questions regarding the *Level One* and *Level Two* trainings. Comparison teachers were also asked if they received training for their reading program and if that training was effective in preparing them to implement the program.

Availability of Program Materials

The ability to implement a program well is dependent on teachers' access to needed materials. Prior to initiation of the study, teachers reported using a variety of materials for reading instruction. Many teachers supplemented their reading program with teacher produced materials. Since *PCI* teachers may have had students using different levels of the program, the surveys asked teachers in both groups to confirm that they have the necessary materials to fully implement their reading programs.

Availability of Teaching Assistants

Many of our participating classes also receive assistance from paraprofessionals, aides, parents, or other adults. Certain components of the *PCI Reading Program* require one-on-one instruction between instructor and student. Therefore, we asked all participating teachers to indicate the types of support they received (such as paraprofessionals or aides) in their classrooms who work with participating students during reading instruction.

Description of Implementation

While the primary focus of this study is on student outcomes, the results need to be understood within the context of classroom implementation. In the following sections, we evaluate the differences between *PCI* and comparison classrooms in order to discern differences that might impact student results. We also include a description of how *PCI* teachers implement the program in their classrooms.

Reading Materials Used

On the background information form included in the consent packets, teachers in both conditions indicated that they used a variety of materials for reading instruction. Therefore, in the surveys we asked *PCI* teachers if they supplemented reading instruction with other materials, and if so, what materials they used. Comparison teachers were asked to describe the reading instructional materials used in their classroom.

Teacher Satisfaction

Teachers in both groups were asked to rate their level of satisfaction with their reading program and whether they would recommend their primary reading program to other teachers of the population of students they teach. Additional questions probed challenges teachers may have experienced, preferences and possible modifications, and plans for future implementation.

Levels of PCI Reading Program Implementation

PCI Education is very interested in learning how closely the implementation complies with their recommendations. Therefore, we collected data to characterize how *PCI* teachers implement the program in the classroom. We investigated how far students progressed through the program during the year and how many students were on *Level One* and *Level Two. PCI* teachers were asked questions about classroom and instructional organization, student assessment, and program bonus materials. Our analysis compares the classroom implementation data against the implementation expectations set in the curriculum and reinforced during the training.

- **Program Level and Student Progress**

  Because the program is designed to allow students to progress at their own pace, researchers tracked how far each student progressed through each level of the program during the school year. At the end of the year, the teacher questionnaire asked teachers to indicate on which level and word each individual student began and ended the year.

- **Lesson Cycle Organization**

  Strict adherence to the Teacher's Guide is specified in the *PCI* curriculum and was reinforced during training. Therefore, researchers tracked how closely teachers adhere to the lesson cycle in their classrooms. The publisher maintains that the *PCI* program can be taught by any adult familiar with the program, and researchers were interested in determining whether other adults in the classroom provided instruction to participating students. *PCI* teachers were asked about the following areas of classroom and instructional organization:

  - How students are organized while instruction is delivered (e.g., one-on-one instruction, group instruction, or independent student work)
  - How regularly each step is completed
  - Who teaches each step in the lesson cycle

- **Student Assessment**

  Assessment is a key component of *PCI*'s mastery-based curriculum and is part of the lesson cycle. To help inform the level of program implementation, data were collected regarding whether teachers were assessing students in the recommended manner.

- **Bonus Materials**

  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. The Building Reading Skills binder includes supplementary materials for *Level One,* but is expected to be used in *Level Two* to prepare students for *Level Three.* Researchers examined the extent to which teachers were using the binder for both levels. For researchers, responses about these materials helped inform our measures of time spent on task. For the publisher, reactions to supplemental materials such as the CD-ROM provide valuable feedback on what the teachers find useful about the program.

Correlation between Implementation Fidelity and Student Outcomes

We also asked *PCI* teachers about the amount of time students received *PCI* instruction during a given week and about the amount of time the instructor taught, other adults taught, and students worked independently. From teachers' answers to these questions, researchers planned to check for a correlation between the amount of instructional time and student achievement outcomes. It can be hypothesized that, as instructional time with the *PCI Reading Program* increases, student achievement would also increase.

Impact of Instruction

Reading Instruction Time

Researchers investigated whether the number of minutes of reading instruction students received was different between conditions. Teachers reported the number of minutes students received reading instruction during a given week. Surveys posed identical questions across seven surveys in order to gain an understanding of variation at different times during the school year.

During Phase 1, researchers had found a significant decrease in the number of minutes of *PCI* instruction reported during the Florida Alternative Assessment (FAA) testing period. Therefore, when the state testing or FAA period began, teachers were asked if instructional practices change and if they supplemented the *PCI* program in any way in order to prepare for standardized testing.

Student Engagement

Surveys also included questions about the level of student engagement with their primary reading materials. Because student engagement is an aspect of the *PCI Reading Program* that is critical to the teachers, it was an important element to measure. We also measured the level of student engagement in the comparison group and the level of engagement with the core pieces of the *PCI* program with the *PCI* teachers.

Denominator (*n*) Counts on Implementation Data

The denominator (*n*) count from the teacher questionnaire on student progress is based on the number of questionnaires returned. The (*n*) counts from survey data are based on the number of teachers who completed the survey from which the data were collected. If a teacher completed the survey but did not provide a valid response for a specific question, the response was coded as missing data.

In some cases, we report survey data for three teacher groups: 1) the comparison teachers, 2) *PCI* teachers whose students were on *Level One* of the program at the time of the surveys, and 3) *PCI* teachers whose students were on *Level Two* of the program at the time of the surveys. Where results are presented separately for *Level One* and *Level Two,* individual teachers may be represented in both groups, as they may have had students on both levels.

Identification of Student Groups and Analysis Plan

Student Groups

Due to the study design and the criteria set for teacher participation, students began Phase 2 with different levels of exposure to the program, including some students who had been part of the initial randomized group in Phase 1. Students are categorized into five groups based on these factors and their data are used in different analyses depending on their group.[8] Table 6 shows the patterns of exposure of the different student groups.

Table 6. Student Groups

Student Group	Phase 1 (2007–2008)		Phase 2 (2008–2009)		
	PCI	Control	*PCI*	Control	Comparison
Group 1	X		X		
Group 2		X	X		
Group 3		X		X	
Group 4	-	-	X		
Group 5	-	-			X

Note. "-" indicates that the student group was not involved in Phase 1.

Figure 1 further illustrates the student group patterns presented in Table 6. Figure 1 is a schematic drawing meant to help the reader distinguish the groups involved in the analysis and the underlying design, including the students' years of exposure to *PCI*. The slopes of the lines are not meant to be interpreted as representing observed gains in student achievement.

[8] As noted earlier, because we focused on the two-year impact of *PCI,* data from some teachers and students were not used in any analysis. It was important to collect data from these students for potential use in for future years of the study.

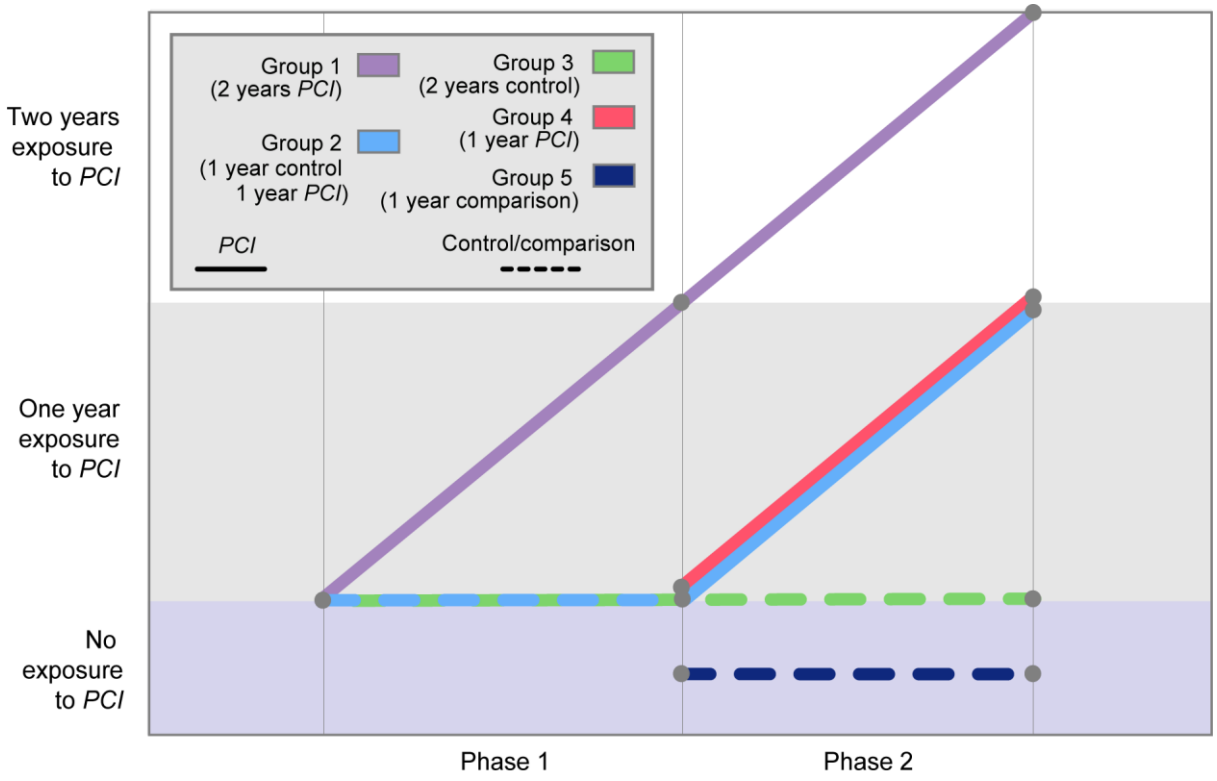

Figure 1. Illustration of Student Groups

Analysis Plan

As mentioned in the Experimental Design section of this report, we use two experimental approaches to obtaining estimates of the impact of the program: a quasi-experimental approach and an "extra-experimental" approach (attributed to Bell and Bradley, 2008). For each approach, we compare different student groups:

1. Quasi-experimental estimate of the two-year impact: This analysis involves students who have used *PCI* for two years (those who were randomized to *PCI* at the start of Phase 1 and continued to use *PCI* in Phase 2) and their teachers, as well as teachers and students who had never received exposure to *PCI*. This analysis compares group 1 to groups 3 plus 5.
2. Extra-experimental estimate of the two-year impact: This analysis involves students who were part of the randomized *PCI* group in Phase 1 and continued to use *PCI* in Phase 2 (and their teachers) and Phase 1 control students who used *PCI* in Phase 2 (and their teachers). This analysis compares group 1 to group 2.

Formation of the Experimental Groups

Characteristics of the *PCI* and Comparison Groups

This section describes the sample that we will use to determine the relationship of *PCI Reading Program* to the measured outcomes.

The program group is composed of teachers who participated in Phase 1 and their students, and teachers of students who received *PCI* instruction during Phase 1 and their students. The

comparison group consists of all other teachers of target students. The comparison group was used in the quasi-experimental analysis. The hope in selecting a comparison group is that the groups are similar in terms of important characteristics such as demographic composition, achievement, and teacher characteristics. However, the groups are never exactly balanced and may be out of balance on important characteristics likely to affect the outcome. In a quasi-experiment, we often have less information compared to a randomized control trial. Furthermore, the loss of teachers and students during the period of program implementation may introduce a bias if, for example, teachers are more likely to drop out of the program group than the comparison group because of the extra burden. Therefore, in this section we inspect the teachers and students and check whether the *PCI* and comparison groups are balanced on important characteristics. (For this accounting, we focus on the data available for Sight Word Assessment results, which we consider the primary outcome measure.) As described in a previous section, there are two sets of analyses: (1) the quasi-experimental analysis of the two-year impact, and (2) the extra-experimental analysis of the two-year impact. The formation of the *PCI* and comparison groups (and thus the sample used) varies by the type of the analysis.

Attrition of Students Due to Assessment Scores

In Phase 1 we considered separately students who had scored a zero on the pretest and those who had scored above zero on the pretest. One rationale for this was that the students with a zero pretest score were potentially consistent with a different kind of student—that they possibly did not understand the task or were not engaged. We analyzed the results for these students separately.

In Phase 2, we had fewer students who had scored a zero on the pretest (and we expect to have even fewer students in this group in future years). The number of students was not big enough for a powerful separate analysis or to cause a serious floor effect. Therefore, students who had scored zero on the pretest were analyzed together with the students who had scored above zero on the pretest in Phase 2. A small number of students hit 20 on the posttest; however, the number of students was not big enough to cause a serious ceiling effect. These students were also included in the analysis. However, students who received zero on both the pre- and posttest were excluded, as we suspect that these students represent a different kind of student from those who scored above a zero on the pre- and posttest. It is possible that these students did not understand and/or had no way of responding to the test, given the way it was administered, and that they would drive the results of the regression.

Number of Units in the Sample and Attrition in the Analysis of the Two-year Impact

The following tables show the reductions in the teacher and student samples from the point at which we defined a starting number of cases to the point when posttests were received. Table 7 shows counts for the quasi-experimental analysis of the two-year impact. Table 8 shows counts for the extra-experimental analysis of the two-year impact. (Numbers in the parentheses show the reductions of the counts in the samples.)

Table 7. Numbers of Units in the Quasi-Experimental Analysis of the Two-year Impact

Event	Comparison				*PCI*			
	No. of teachers		No. of students		No. of teachers		No. of students	
Starting sample	13		64		12		33	
(Excluded in matching process[a])	(0)		(5)		(1)		(5)	
Retained after matching	13		59		11		29	
Posttest outcome	**SW**	**PH**	**SW**	**PH**	**SW**	**PH**	**SW**	**PH**
(Removed because of perfect score at pretest [i.e., a score of 20])	(0)	(0)	(2)	(0)	(0)	(0)	(1)	(0)
(Removed because pretest is missing)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)
(Removed because of missing roster information)	(0)	(0)	(1)	(1)	(0)	(0)	(0)	(0)
Students considered for analysis	13	13	56	58	11	11	28	29
(Loss before/at posttest)	(0)	(0)	(5)	(4)	(0)	(0)	(2)	(2)
(Removed because of pretest and posttest scores equal zero)	(0)	(0)	(3)	(0)	(0)	(0)	(0)	(0)
Final count of units with pretest and posttest	**13**	**13**	**48**	**54**	**11**	**11**	**26**	**27**

[a] Treatment cases were excluded if they were missing information on covariates required to carry out the matching strategy. Comparison cases were excluded if they lay outside the range of propensity scores for the treatment cases.

Table 8. Numbers of Units in the Extra-experimental Analysis of the Two-year Impact

Event	Comparison				*PCI*			
	No. of teachers		No. of students		No. of teachers		No. of students	
Starting sample	8		24		12		33	
(Loss after taking out students who were not randomized to conditions at the start of Phase 1)	(2)		(8)		(0)		(0)	
Extra-experimental starting sample[a]	6		16		12		33	
Posttest outcome	**SW**	**PH**	**SW**	**PH**	**SW**	**PH**	**SW**	**PH**
(Removed because of perfect score at pretest [i.e., a score of 20])	(1)	(0)	(1)	(0)	(2)	(0)	(2)	(0)
(Removed because pretest is missing)	(2)	(2)	(2)	(2)	(1)	(1)	(1)	(1)
(Removed because of missing roster information)	(0)	(0)	(0)	(0)	(2)	(3)	(2)	(3)
Students considered for analysis	5	5	13	14	11	11	28	29
(Loss before/at posttest)	(1)	(1)	(1)	(1)	(0)	(1)	(0)	(1)
Final count of units with pretest and posttest	**5**	**5**	**12**	**13**	**11**	**11**	**28**	**28**

[a] The extra-experimental estimate is based only on those students who were randomized at the start of Phase 1.

Balance Check for Characteristics of Teachers and Students

Table 9 shows some of the background characteristics of all qualified teachers and students (given in the last row of Table 7) used in the quasi-experimental analysis of the two-year impact for the sight word assessment.[9]

We see that there is balance between conditions in teachers' average years of teaching experience, in student characteristics (number autistic, gender, ethnicity, social economical status, and sight word pretest). We also see that students are not balanced in terms of phonological pretest. Because the phonological pretest was greatly imbalanced between conditions, we excluded that covariate and did not perform an analysis of the phonological outcome. We reasoned that a statistical adjustment would not fix the problem of having lack of support for that covariate. (Given imbalance on the phonological pretest, and our decision not to statically adjust for it, we stress the need for comparing the results of the quasi-experimental analyses to those from the

[9] In a later section on attrition, we briefly discuss the equivalence tests for the analytic sample used for the extra-experimental analysis

extra-experimental analyses; the results of the latter of these analyses are not sensitive to the effects of imbalance on the phonological pretest.)

Table 9. Characteristics of the Teachers and Students

	Comparison group	*PCI* group	Is the imbalance significant?
Teachers:			
Fewer than four years Special Education teaching experience	9 (75.00%)	8 (72.72%)	No
Students:			
Black	9 (16.07%)	5 (17.86%)	No
Verbal	4 (85.71%)	24 (92.30%)	No
Male	37 (66.07%)	16 (57.14%)	No
National School Lunch program	32 (57.14%)	21 (75.00%)	No
Autistic	7 (12.50%)	8 (28.00%)	No
Mean sight word pretest	8.05	6.07	No
Mean phonological pretest	8.64	1.32	Yes

Descriptive Characteristics of Teachers and Classroom Descriptions

Here we report additional characteristics of for all study teachers and classrooms. Data presented in the following sections were collected through teacher surveys and are included for descriptive purposes only.

Teacher Background

At the beginning of the 2008-09 academic year, teachers responded to survey questions regarding their teaching experience and educational background. Provided in Table 10 and Table 11 are the responses regarding years of teaching experience for each condition, as well as the number of years of experience each condition has teaching Special Education. All of the comparison teachers and all but one of the *PCI* teachers have spent their entire teaching careers in Special Education. More of the *PCI* teachers fall into the “experienced” category (16+ years) than among the comparison group teachers (52% compared to 46%).

Table 10. Overall Years of Teaching Experience

	0–3	4–6	7–15	16+
Comparison (*n* = 13)	3 (23.1%)	0 (0.0%)	3 (23.1%)	6 (46.2%)
***PCI* (*n* = 21)**	3 (14.3%)	2 (9.5%)	5 (23.8%)	11 (52.4%)

Note. We are missing data from one comparison teacher (7.7%)

Table 11. Years of Teaching Special Education

	0–3	4–6	7–15	16+
Comparison (*n* = 13)	3 (23.1%)	0 (0.0%)	3 (23.1%)	6 (46.2%)
***PCI* (*n* = 21)**	4 (19.1%)	2 (9.5%)	4 (19.1%)	11 (52.4%)

Note. We are missing data from one comparison teacher (7.7%)

All but one teacher in each assignment group held a regular/standard teaching certificate. The remaining two teachers held temporary certificates at the time of the survey. All teachers, except for one comparison teacher, reported that they were licensed to teach Special Education or had a teaching certificate in Special Education.

Table 12. Teacher Credentialing and Certification

	Regular/ standard	Temporary	NBPTS[a]	Specific certificates for teaching bilingual, multicultural, or limited English	None	Other[b]
Comparison (*n* = 13)	12 (92.3%)	1 (7.7%)	2 (15.4%)	3 (23.1%)	0 (0.0%)	1 (7.7%)
***PCI* (*n* = 21)**	20 (95.2%)	1 (4.76%)	2 (9.5%)	1 (4.8%)	0 (0.0%)	3 (14.3%)

[a] NBPTS: National Board for Professional Teaching Standards.

[b] Other included ESOL K-12, Autism endorsement, reading endorsement, Language Arts grades 5-9, Speech/Language, CCC/SLP.

Note. Because teachers could select more than one category, totals may exceed 100%.

A larger percentage of teachers in the comparison group had obtained a Master's degree (8 out of 13 or 62%) than among teachers in the *PCI* group (8 out of 21 or 38%).

Table 13. Highest Level of Education Completed

	Bachelor's degree	Master's degree	Education specialist or professional diploma	Doctorate or first professional degree
Comparison (*n* =13)	4 (30.8%)	8 (61.5%)	1 (7.7%)	0 (0.0%)
***PCI* (*n* = 21)**	11 (52.4%)	8 (38.1%)	2 (9.5%)	0 (0.0%)

Classroom Description

Within both groups there was a range of classroom classifications. However, there were more ASD and TMH classrooms in the *PCI* group (5 out of 21 or 25% and 7 out of 21 or 35%, respectively) than in the comparison group (1 out of 12 or 8% and 3 out of 8 or 25%, respectively). One teacher in the *PCI* group reported that her classroom included EMH and TMH students (her classroom also included PMH students, but these students were not included in the study). Another *PCI* teacher reported that his classroom classification was "orthopedically impaired with mental disabilities." Both of these teachers' responses are represented in the "other" category in Table 14.

Table 14. Classroom Description (of Participating Students)

	Intellectual disabilities at the supported level	EMH	TMH	PMH	ASD	Varying exceptionalities	Other
Comparison (*n* = 12)	4 (33.3%)	2 (16.7%)	3 (25.0%)	0 (0.0%)	1 (8.3%)	2 (16.7%)	0 (0.0%)
***PCI* (*n* = 20)**	6 (30.0%)	0 (0.0%)	7 (35.0%)	0 (0.0%)	5 (25.0%)	0 (0.0%)	2 (10.0%)

Note. We are missing data from two teachers (one comparison and one *PCI*).

Teacher and Student Attrition

Teacher Attrition

By the end of the study year, a total of four teachers had left the study after consenting to participate. Two comparison teachers from M-DCPS decided to leave the study citing a decrease in support staff. One of these teachers also said that she did not believe her students would be able to take the pre-assessments, due to their ability level. Two other teachers from

M-DCPS (one comparison and one *PCI*) were unable to obtain parental consent from their students.[10] Both teachers said that they sent the consent forms home with their students several times but did not receive returned signed forms. While neither teacher formally withdrew from the study, we are counting them as attrition because we were unable to receive any student data. Current participant totals are reflected in the tables and text above.

Student Attrition for Extra-Experimental Analysis

The validity of the results depends on ruling out plausible alternative explanations for the differences that we observe. In particular, for the extra-experimental outcome, we are interested in whether, over time, the attrition of students from the study may have led to groups that are non-equivalent in systematic ways that may lead to misinterpretation of the program effect.

The rate of attrition for students who were part of the original experimental sample was large. In Phase 1, for all rosters received, if we consider students with pretests, we had 76 control students and 77 treatment students. In the extra-experimental analyses in Phase 2 that we have reported above, we have 37 students with posttests at the end of Phase 2 (11 students received exposure to *PCI* for the first time in Phase 2, and 26 received exposure to *PCI* for the first time in Phase 1). This represents an overall attrition rate of 86% for students originally in the control group, and 66% for the students who received *PCI* in Phase 1. The rate of attrition is different for the groups originally in the two conditions; however, the difference in pretests between those who remain versus those who leave the study is not statistically significant. This suggests that differential attrition will have a limited biasing effect.

We were also interested in whether the groups in the two conditions in the analytic sample used to obtain the two-year extra-experimental impact estimate were statistically equivalent on a set of covariates. We examined whether the two groups were imbalanced on the following covariates measured at the start of Phase 1: sight word pretest, phonological pretest, National School Lunch Program status, and grade level. The groups were balanced on these covariates. Although there may be additional covariates that are imbalanced between the study groups, for this set of covariates and, very importantly, for the pretest, we see that the balance that was established through randomization is maintained at the end of Phase 2.

Despite the fact that the reduced sample passes the tests described above, the level of attrition is large when we consider the analytic sample used for the extra-experimental analysis. This reinforces the need to have both quasi- and extra-experimental results to compare when discussing two-year impacts on the sight word outcome.

An additional point to consider regarding attrition is that the power to detect the two-year impact of *PCI* on phonological outcomes may be very limited, given the available sample of cases. The impact of *PCI* on sight word performance is large and is therefore robust to the loss of cases that we have experienced; if the impact of *PCI* on the phonological outcome is small then not seeing an effect may be the result of having insufficient power to do so, rather than due to the absence of the effect. To understand the impact of *PCI* on the phonological outcome will require

[10] The district requires two forms of parental consent for the release of student data. One form was provided by the district and the other was a letter to parents/guardians written by researchers describing the study and expectations. Both forms ask parents to consent to allow their child's assessment/demographic data to be used in the analysis of this study. The letter to parents was also made available in Spanish and Creole translations. The forms were handed out to teachers at the training and delivered by district support staff members to those who could not attend the training. Upon receiving signed parental consent forms, teachers were asked to fax/mail the forms to researchers. Researchers then mailed all signed district consent forms to our district contact, who provided student data to us for all students for whom they had signed consent forms. Teachers were not able to return assessment data to researchers unless they received parental consent forms.

replication studies for the impact to be estimated with sufficient power. Attrition has also limited our ability to analyze the moderating effects of certain covariates. Here too we need additional studies with larger samples that include subgroups of students across which we hypothesize *PCI* to have differential effects.

Statistical Equations and Reporting on the Impact of *PCI Reading Program*

Setting Up the Statistical Equation[11]

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 outputs of this process are estimates of effects as well as a measure of the level of confidence we can have that the estimate is close to its true value.

Program Impact

A basic question for the experiment was whether, following the intervention, students in *PCI* classrooms had higher Reading scores than those in comparison classrooms, as measured by the sight word and phonological outcomes. Answering this is not as simple as comparing the averages of the two groups. It is also essential that we understand how much confidence we can have that there really is a difference between the two groups, given our estimate of the difference in outcomes between the program and comparison groups. To appropriately estimate this difference in the quasi-experimental analysis, our equation contains a term for *PCI* as well as terms for other important factors including the student pretest score, the propensity score, and other covariates. The student's prior score is, of course, an important factor in estimating his or her outcome score. By including the pretest as a term in the statistical equation, we are able to improve the precision of this estimate because it helps to explain much of the variance in the outcome and makes it easier to isolate the difference associated with the program. A second goal of including these covariates is to control for systematic differences between conditions on these covariates—imbalances that could produce selection bias in the estimate of

[11] 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 the 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 program and control groups. In the current study, the statistical equation used to obtain the extra-experimental estimate takes this form. To obtain the quasi-experimental estimates of impact we use statistical equations that take the same form as the one described above except that the propensity score and possibly other covariates are included in the equations, thereby allowing us to estimate the effect of *PCI* conditional on the covariates. In other words, we estimate the program effect while holding constant the values of the covariates across the program and control conditions.

the effect of *PCI*. We also have to account for the fact that students are 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 factor this dependency into our equation or else our confidence levels about the results will be artificially high. (The statistical equation used to obtain the extra-experimental estimate takes a simpler form. It includes the treatment indicator, the pretest and a term to account for clustering of students in teachers. Because the samples are statistically equivalent as a result of the randomization at the start of Phase 1, it is unnecessary to include additional covariates to adjust for imbalance due to systematic differences between conditions.)

Covariates and Moderators at the Student and Teacher Level

When we estimate differential impacts beyond just the average impact, we also include in the equation other variables (called covariates) associated with characteristics of teachers or students, which we expect to make a difference in the outcomes. For example, as was described above, we add the pretest score into nearly 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 comparison 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.

Teacher Level Outcomes and Potential Mediators

We are also interested in measurable characteristics of the schools, classrooms, teacher behavior, or beliefs, student activity that can be observed or ascertained during the experiment. Unlike the moderators, these are not pre-existing characteristics such as pretest score or years of experience. In fact, they can sometimes be shown to be the result of introducing the new program. These factors are called "potential mediators" because they can be understood as standing between the introduction of the program and the outcome. In other words, they can be understood as a way that the program has its impact (or, in some cases, as a way that the program is suppressed). We conceptualize mediators as being triggered by the introduction of the program. For example, the program could cause changes in instructional practices and we would want to show whether a program's effect on student outcomes is, in whole or in part, due to these changes in instruction. Instructional practice is a mediating variable if we can show that at least some of the program's effect works through changes of instruction. In effect this mediator analysis sets up two stages of equations, from *PCI* to mediator (and from the mediator to student achievement). We can compare the effect of the program as it works along this path to the direct effect from program to outcome.

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 population. We can also treat as fixed the levels of a factor that we have sampled from a population. This reduces uncertainty because it does not figure in the variation in the outcome that we expect from a re-sampling of units from the population, but it also keeps us from generalizing the results to the population from which we sampled.

We usually treat the units that are assigned in their entirety to the program 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 such units from the same population.[12] 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 degree 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 we see among students that are not accounted for by all the fixed effects and other random effects in our statistical equation.

The statistical equations that are used to test moderator effects often include a random term for the slope of the moderator. That is, we allow the quantity that relates the expected change in the outcome for each unit increase in the potential moderator to vary from one upper-level unit to the next (usually from one unit of assignment to the next). This allows the uncertainty in the estimate of the moderator effect to reflect not only variation due to the potential re-sampling of students, but also the re-sampling of the higher-level units. In this experiment, for example, we model the pretest slopes as random across teachers, so that we can measure the moderating effect of pretest under uncertainty due to potential re-sampling of students and teachers. In the current study we treat teachers and students as random factors. We do not model the effects of individual schools. Program and comparison cases were for the most part from different schools. By modeling effects at the student- and teacher-levels, we adjust for the differences between conditions in the school-means of these variables.

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 computations will provide an explanation of these choices in table notes where necessary for technical review.

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 comparison groups into a standardized effect size by dividing the average group difference by the amount of variability in the outcome. The

[12] 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 is fair to ask how much change in outcomes we can expect from this substitution.

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 the pretest score and other fixed effects, since that adjustment provides a more precise estimate of the effect by compensating for average differences on these covariates between the program and comparison groups. Theoretically, since with many replications of the experiment these differences would wash out, we would expect the adjusted effect size on average to be closer to the true value. For quasi-experiments we calculate the adjusted effect size in the way that is described above; however, we also make adjustment for imbalance on covariates that could lead to bias. For instance, in this study, we use the effect estimate from a model that conditions on the propensity score.

Estimates

We provide estimates to approximate the actual effect size. Any experiment is limited to the smaller sample of students, teachers, and schools that represents 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 comparison 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 comparison to the program group (while holding constant the other covariates in the model).

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 an absolute 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 it actually hasn’t. This mistake is also known as a “false-positive” conclusion. Thus a *p* value of .1 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

Implementation Results

In this section, we provide a description of the implementation of the intervention and comparison groups to inform the readers' interpretation of student outcomes. Within the *PCI* group, we also provide an extensive description of the level of implementation of the *PCI Reading Program* to examine whether expectations set in the curriculum and reinforced during the training were met. Data for this section were obtained through surveys, teacher questionnaires on student progress, classroom observations, informal teacher interviews, and formal interviews with principals.

Conditions for Implementation

Here we provide a description of the conditions under which implementation in each assignment group took place. We report findings on the amount of training and level of training effectiveness as well as the availability of materials for reading instruction and the availability of teaching assistants.

Training

All *PCI* teachers were offered training in the implementation of the *PCI* program—18 out of 21 (86%) attended the training at the beginning of Phase 2. Of the remaining three teachers, one attended the training during Phase 1 and the other two were offered training by a district support specialist, but did not receive training. Five of the 13 (39%) comparison teachers reported receiving training for their current reading program, two of whom (15%) had been trained within the previous year. Four of the five comparison teachers who received any training in their reading program reported it to be moderately effective, with the remaining teacher reporting a neutral opinion.

Figure 2 and Figure 3 present *PCI* teacher responses to survey questions regarding the effectiveness of the *PCI* training in preparing them to implement specific components of each *PCI* program level. The majority of the teachers who had attended the full-day training, and who were teaching the respective program levels, indicated that the training was effective or very effective in preparing them to implement those components.[13]

[13] Teachers who did not attend the training or were not using the respective program levels were not asked for their opinions regarding training.

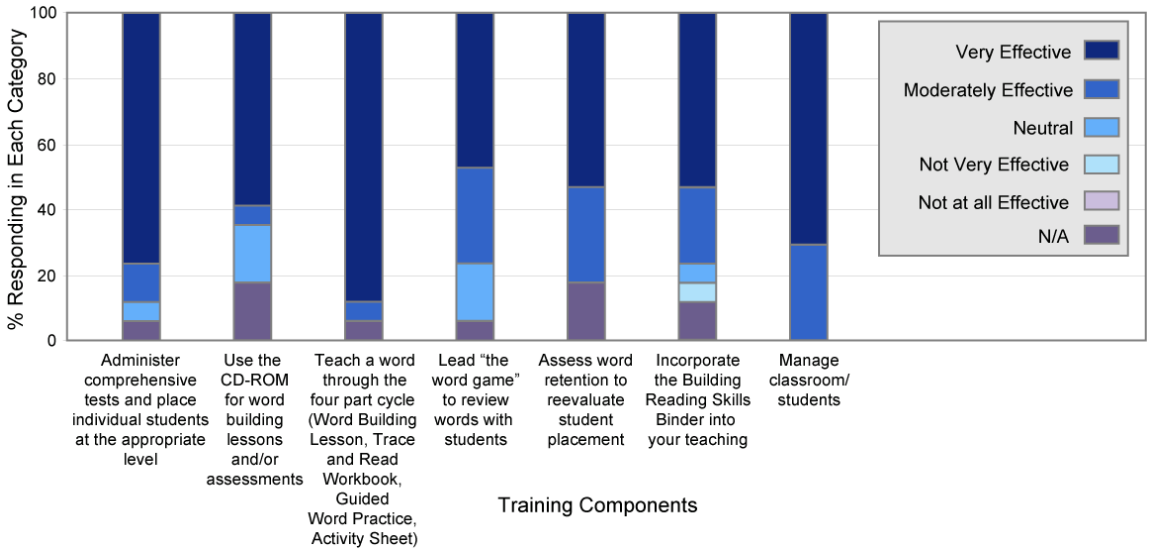

Figure 2. Effectiveness of *PCI Level One* Training by Task Area

Note. (*n* = 17)

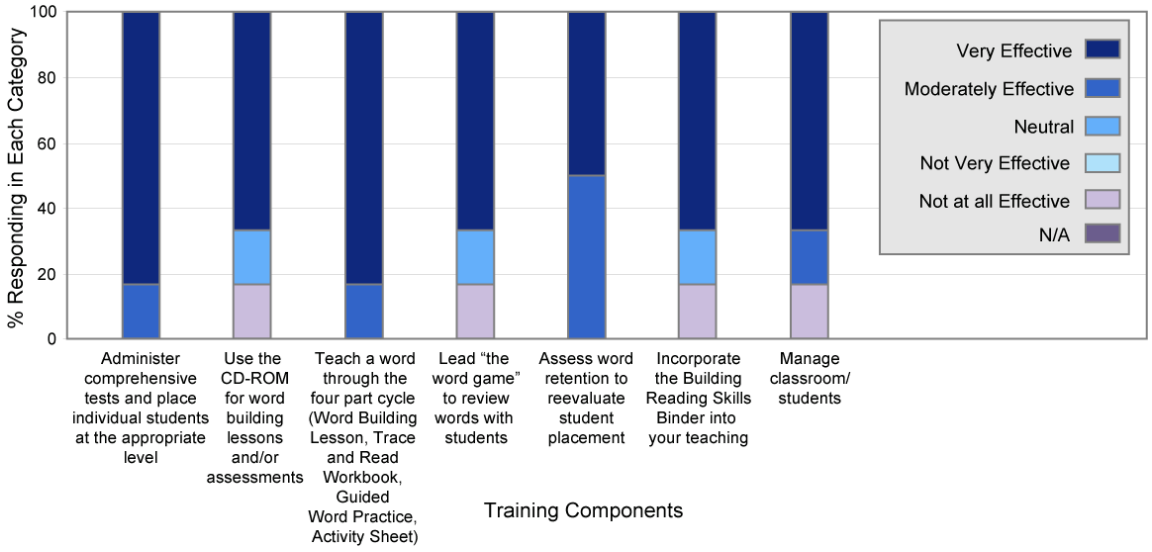

Figure 3. Effectiveness of *PCI Level Two* Training by Task Area

Note. (*n* = 6)

Availability of Materials

The *PCI* teachers who attended the initial training in October and November 2008 received the *PCI* materials at that time.[14] When they were surveyed in January 2009, 100% of *PCI* teachers reported to have all the materials needed to fully implement the *PCI* program. Only 10 out of 13 comparison teachers (77%) reported having all the materials needed to fully implement their reading programs.

Availability of Teaching Assistants

All comparison teachers reporting having additional assistance in their classrooms, while 89% (17 out of 19) of the *PCI* teachers reporting having additional assistance. Within the *PCI* group, 11 out of 19 teachers (58%) had a general paraprofessional in their classroom. Throughout the seven surveys in which we asked teachers to account for the minutes of program instruction provided by other adults in the room; all but one teacher reported minutes for other adults.

When the researcher conducted classroom observations, paraprofessionals or one-to-one aides were present in all six of the observed *PCI* classrooms. In 4 out of the 6 classrooms, the paraprofessional was implementing *PCI* one-on-one with students. In the other two classrooms, the teacher was implementing the program one-on-one as the paraprofessional managed the other students. Five out of the 9 comparison classrooms observed had paraprofessionals or one-to-one aides present. However, in all nine classes, the teacher was delivering the primary instruction to the whole class or small groups, while the other adults present worked with individual students.

Table 15. Classroom Support for Reading Instruction

	Co-teacher	Assistant for clerical tasks	General para-professional	Tutor for individual students	Aide or non-professional	Trained specialist for small groups	No professionals or para-professionals	Other[a]
Comparison (*n* = 13)	1 (7.7%)	0 (0.0%)	5 (38.5%)	2 (15.4%)	7 (53.9%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
***PCI* (*n* = 19)**	1 (5.26%)	0 (0.0%)	11 (57.9%)	4 (21.1%)	4 (21.1%)	1 (5.3%)	2 (10.5%)	2 (10.5%)

[a] Other included an intern and a teaching assistant.

Note. Because teachers could select more than one category, totals may exceed 100%.

Summary

Overall implementation conditions across both groups were comparable and generally good for implementing the *PCI Reading Program*. Fewer than 40% of the comparison teachers had been trained in their existing reading program, and those that were trained reported moderate effectiveness of the training, whereas *PCI* provided training to more than 86% of the teachers

[14] As previously explained, the *Level Two* kits did not arrive in time for the training In BPS. However, teachers received the kits the following week.

using the program and many of these teachers felt that the training was very effective. Of the *PCI* teachers, 100% reporting having all of the materials needed to implement the program. All teachers in the comparison group, and all but two teachers in the *PCI* group, had other adults in the classroom to provide additional assistance.

Description of Implementation

Here we present our findings regarding classroom implementation. We describe the reading instructional materials used and the level of teacher satisfaction with those materials for both assignment groups. We also provide further information on how teachers who used *PCI* implemented the program in their classroom.

Reading Materials Used

PCI Classrooms

A majority of the teachers began using *PCI* in October or November 2008 and continued through to the end of the year in May or June 2009. By December 2008, 17 out of 21 *PCI* teachers (85%) had begun instruction with the *PCI* program. By January, all but one teacher had begun instruction with the program. This teacher began using the program by early March. During classroom observations, the researcher observed clues, such as *PCI* materials still wrapped in plastic, indicating that a few of the *PCI* teachers had spent very little time using the program.

By March 2009, a majority of the teachers reported supplementing reading instruction with other materials. Thirteen out of 20 teachers (65%) using *Level One* and 3 out of 4 of teachers (75%) using *Level 2* reported using other materials to supplement reading instruction. By April 2009, all *PCI* teachers reported using supplemental materials. A variety of materials were reportedly being used, including Brigance, EdMark Functional Word series, *PCI* Safety Signs, The Letter People, Dolch words, News-2-You, Leap Pad, Starfall, teacher-made worksheets, and required district materials. During classroom observations, the researcher observed materials other than *PCI* being used in 4 out of 6 classrooms.

Comparison Classrooms

Comparison teachers reported using a variety of materials for reading instruction. Common curriculum materials included Houghton Mifflin (four teachers), EdMark (three teachers), Sight Word Readers series from Scholastic (two teachers), and Scott Foresman Reading (two teachers). Four additional teachers reported using teacher made materials. The researcher verified the use these materials during classroom observations. In addition to the curricula listed above, the researcher observed teachers using story books, alphabet games, Triumphs and Treasures textbooks, and a variety of materials to teach functional skills (such as reading a calendar or menu).

Teacher Satisfaction

At the end of the academic year, teachers in both assignment groups were asked for opinions regarding their current curriculum. As shown in Figure 4, 12 out of 19 in *Level One* teachers (63%) and 4 out of 6 in *Level* Two teachers (67%) reported that they were very satisfied with the program, while only one comparison teacher reported that opinion of their reading program. Not one *PCI* teacher reported dissatisfaction with the program.

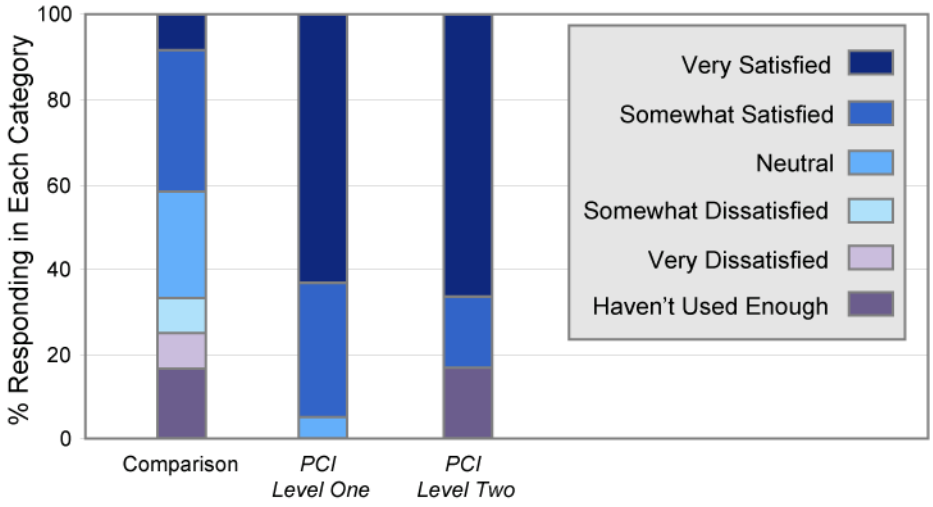

Figure 4. Teacher Opinion of Reading Program: Comparison Versus *PCI*

Note. Comparison (*n* = 12); *PCI Level One* (*n* = 9); *PCI Level Two* (*n* = 6)

Figure 5 elaborates on how teachers rated the Building Reading Skills Binder and the CD-ROM, optional or supplemental components of the *PCI* program. For each component, most teachers who used each piece of the program enough to form an opinion were generally satisfied with these aspects of the program.

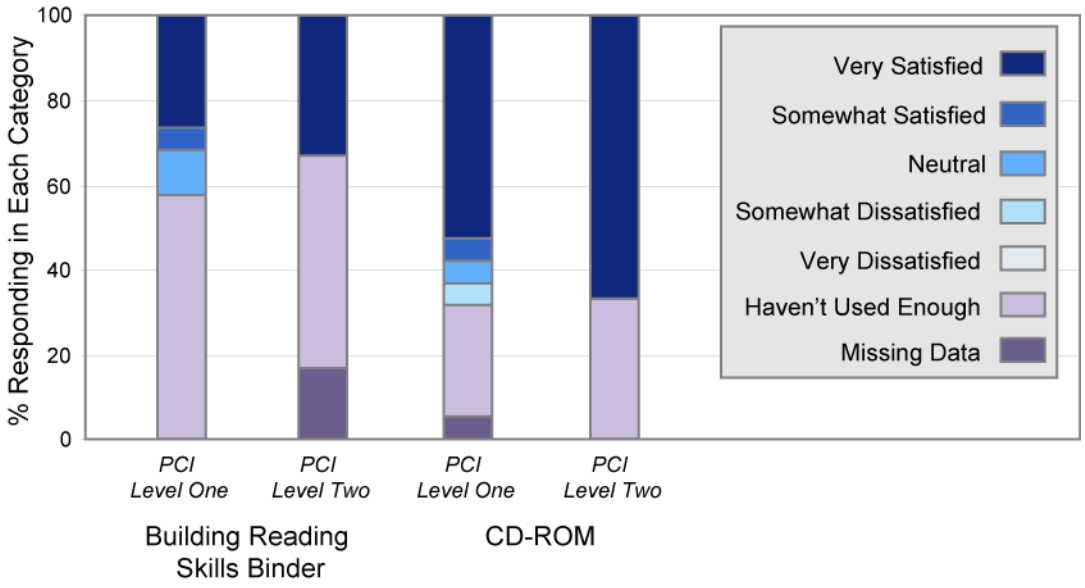

Figure 5. Teacher Satisfaction with Optional/Supplemental *PCI* Components

Note. *PCI Level One* (*n* = 19); *PCI Level Two* (*n* = 6)

Researchers also asked teachers in both groups whether they would recommend their reading program to other teachers. Of the 19 teachers using *Level* One, 18 (95%) reported that they would recommend the program. Of the six teachers using *Level Two,* four (67%) reported that they would recommend the program. In contrast, three comparison teachers (25%) said that would recommend their current program to teachers of this population.

Table 16. Would you recommend reading program to other teachers?

	Yes	No	I don't know
Comparison (*n* = 12)	3 (25.0%)	1 (8.3%)	8 (66.7%)
***PCI Level One* (*n* =19)**	18 (94.7%)	0 (0.0%)	1 (5.3%)
***PCI Level Two* (*n* = 6)**	4 (66.7%)	0 (0.0%)	2 (33.3%)

In response to open-ended surveys questions on what teachers found useful and difficult about their programs, respondents in both groups provided many descriptive comments. Four common themes that emerged from teachers about the usefulness of *PCI*:

- the repetition and review within the program,
- that students were engaged
- that it fulfilled an important need for this population of students
- that student progress was visible throughout the year

The primary difficulty reported by *PCI* teachers was finding the time for individualized instruction (since much of the program is administered one-on-one). Other themes included preparing copies of materials for their students and having to "go back-and-forth" between materials in order to complete the lesson cycle.

Within the comparison group, teachers commented both positively and negatively about the variety of materials they used in their classrooms. Some teachers liked that the variety allowed them to differentiate instruction for meeting the needs of individual students, while others were critical of the lack of a standardized curriculum and said it was too time consuming to find resources that are not even necessarily made for their population of students.

PCI Reading Program Levels of Implementation

Here we examine how far *PCI* students progressed through the program during the year, how teachers organized instruction, how students performed on and how teachers used the program assessments, how teachers used the bonus program materials, and whether teachers would continue to use the program.

Program Level and Student Progress

By the end of the academic year, 83 students were on *Level One* of the program and six students were on *Level Two*. Twenty teachers reported using *Level One* with their students and

eight teachers reported using *Level Two*.[15] As displayed in Figure 6, almost half of the students that began with *Level One* ended the school year on *Level One* words 1-20. Figure 5 shows that no student progressed beyond word 220 in *Level Two*.

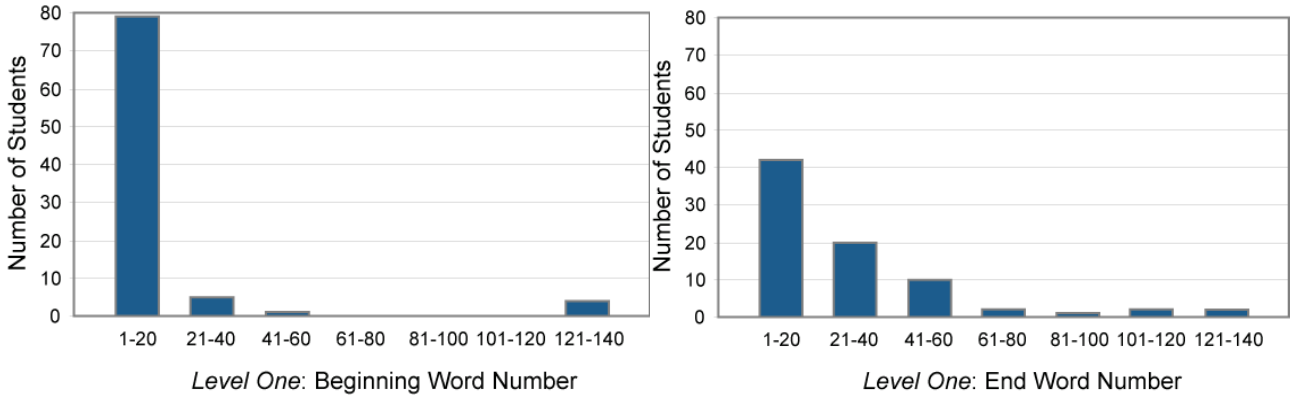

Figure 6. Student Progress: *Level One*

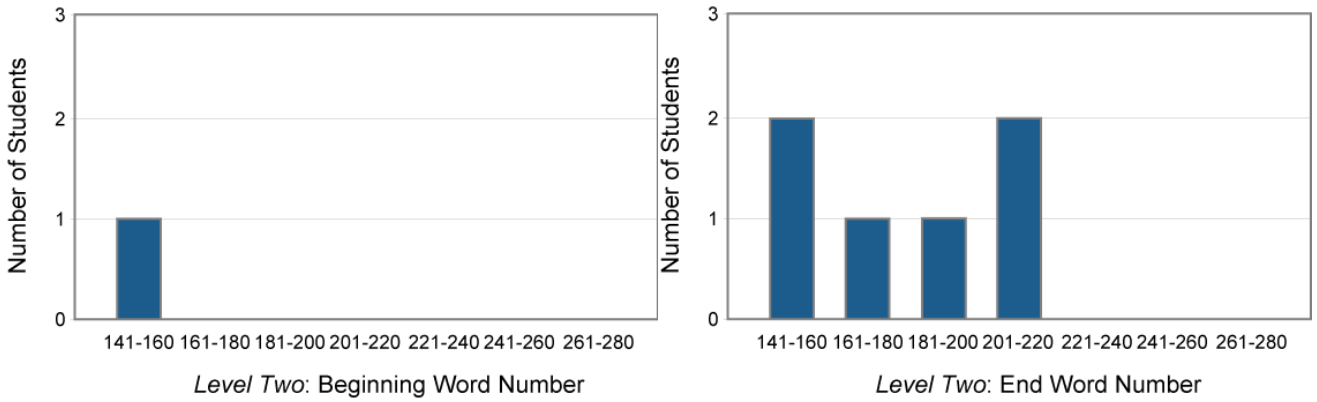

Figure 7. Student Progress: *Level Two*

Lesson Cycle

The survey data in the appendix show the extent to which teachers said they followed the prescribed lesson cycle. This question was asked three times over the course of implementation

[15] Teachers reported student progress at the end of the school year. Teacher use of the program levels was reported on surveys throughout the academic year. We are missing progress data for some students, which may account for the inconsistent number of teachers and students using *Level Two.*

and the tables display the averages of those responses. The responses were calculated and reported separately for *Level One* and *Level Two*.

In general, teachers followed the first step in the lesson cycle, but only 24% of *Level One* teachers and 13% of *Level Two* teachers reported always reviewing the word being studied with "The Word Game." Teachers using both levels reported that most of the instruction was organized either one-on-one or in groups, with the Trace and Read Workbooks and activity sheets also done as independent work. Instruction was divided between the teacher and other adults in the classroom.

Student Assessment in Lesson Cycle

In step 4 of the lesson cycle, teachers administer a posttest which, for *Level One,* contains the 15 most recently learned words and five previously learned words chosen at random and, for *Level Two,* contains the 20 most recently learned words and five to ten previously learned words chosen at random. According to the program's Teacher's Guide, "any word not mastered on a posttest should be reviewed by repeating the appropriate Word Building Lesson, Trace and Read Workbook page, and Activity Sheet." In February 2009, teachers were asked how students usually perform on the posttest and what they usually do if a student misses a word on the posttest. Table 17 shows that, for *Level One,* fewer than half of the teachers reported that their students master a majority of the words. Two teachers who selected "Other" indicated that students will master most of the words, but if they have problems with certain words, then the problem remains. Two other teachers said their students had not yet mastered 20 words at the time of the survey, and one other teacher said that performance on the assessment depends on the individual student. A majority of the *Level Two* teachers reported that students master the words (the one teacher who selected "Other" had not administered a posttest at the time of the survey).

Table 17. Student Performance on Assessment in Lesson Cycle

	Students are usually able to master the majority of the 20 words	Students usually master the most recent 5 (or 10) words, but have difficulty retaining words taught in previous lessons	Students have difficulty with both old and new words	Other
***Level One* (*n* = 18)**	8 (44.4%)	3 (16.7%)	2 (11.1%)	5 (27.8%)
***Level Two* (*n* = 5)**	4 (80.0%)	0 (0.0%)	0 (0.0%)	1 (20.0%)

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

Table 18 shows that 12 out of the 18 *Level One* teachers (67%) adhered to the Teacher's Guide and would re-teach the word lesson if a student missed a word on the posttest. However, most *Level Two* teachers (80%) indicated that they would only do a quick review and move on. No teacher in either group reported that they skipped the review and moved on to keep pace with other students.

Table 18. What do you do when a student misses a word on the posttest?

	Go back and re-teach the word lesson	Do a quick review and move on	Skip the review and move to keep pace with other students	Other
***Level One* (*n* = 18)**	12 (66.7%)	8 (44.4%)	0 (0.0%)	2 (11.1%)
***Level Two* (*n* = 5)**	2 (40.0%)	4 (80%)	0 (0.0%)	0 (0.0%)

Note. Because teachers could select more than one category, totals may exceed 100%.

Bonus Materials

The Activity Sheets are part of the lesson cycle but, because they are reproducible, teachers have the *option* of sending them as homework. As of March 2009, of the 20 *Level One* teachers who responded, 17 (85%) reported using the resource for in-class exercises, six (30%) sent Activity Sheets as homework, and one teacher (5%) reported never having used the resource. Of the four *Level Two* teachers who responded, all four reported using the resources for in-class exercises, and two (50%) sent Activity Sheets as homework.[16]

During the same survey, teachers were asked whether they used print materials or the CD-ROM for word building lessons and/or assessments. Of the 20 *Level One* teachers who responded, three (15%) teacher reported using only the CD-ROM for these tasks, ten (50%) reported using only the print materials, and seven (35%) used both the CD-ROM and print materials at some point. Of the four *Level Two* teachers, two (50%) reported using only the CD-ROM for these tasks, and three (50%) reported using both the CD-ROM and print materials.

In *Level One,* the Building Reading Skills Binder is an optional supplement provided by the program to address students with additional needs. This resource is available for teachers who have students that may need additional support, including help with phonics. By March 2009, 12 out of 20 teachers (60%) had utilized this resource. Eleven of the 20 *Level One* teachers (55%) reported using unit 1, five teachers (25%) reported using unit 3, and four teachers (20%) reported using units 2, 4 and 5 of the binder. While the binder is optional in *Level One,* it is expected to be used in *Level Two* to prepare students for *Level Three.* However, of the four teachers on *Level Two* in March, none reported using the binder with their students.[17]

Continued Use

As they did in Year 1, *PCI* teachers in Year 2 expressed satisfaction with the program. A majority also reported that would continue to use the program. In the final survey of the year, we asked *PCI* teachers if they planned to continue using the program once the research study was over. Seventeen out of the 19 *Level One* teachers who responded (90%) and 4 out of the 6

[16] Because teachers could select more than one category, totals may exceed 100%.

[17] In May 2009 teachers were asked about their opinion of the Building Reading Skills Binder and one *Level Two* teachers provided an opinion, indicating that he/she had used the binder.

Level Two teachers (67%) planned to continue to use the program. No teacher using either level said they planned to discontinue using the program.

Table 19. Do you believe you will continue teaching the *PCI Reading Program* 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
***Level One* (*n* = 19)**	11 (57.9%)	6 (31.6%)	0 (0.0%)	0 (0.0%)	2 (10.5%)
***Level Two* (*n* = 6)**	2 (33.3%)	2 (33.3%)	0 (0.0%)	0 (0.0%)	2 (33.3%)

Correlation between Implementation Fidelity and Student Outcomes

Researchers had planned to analyze whether the minutes of *PCI* instructional time are correlated with student achievement outcomes. However, because we did not collect the minutes of instruction that each individual student received, we are not able to differentiate high-performing students from low-performing students within the same class in terms of the amount of instruction they received. Without student-level time data, we do not believe that the test would yield robust results.

Summary

Comparison teachers reported using a variety of materials for reading instruction, with no standard curricula available. As in Phase 1, during trainings and observations, and on surveys, *PCI* teachers continued to show enthusiasm for and satisfaction with the program. A few teachers experienced a delayed start with the program and all teachers had supplemented reading instruction with other materials. Nearly half of the students who began on *Level One* remained on words 1-20 by the end of the year. Teachers generally followed the lesson cycle and used the additional supplemental materials; however, *Level Two* teachers had not yet used the Building Reading Skills Binder as intended by the publisher. Nearly all teachers said they would continue using the program.

Principals and teachers in both groups expressed a concern about the lack of curricula available for supported level classrooms. *PCI* teachers were encouraged that the *PCI Reading Program* helps fulfill that need.

Impact on Instruction

Here we report the impact of *PCI* on instruction. We report the average minutes of reading instruction and level of student engagement within both assignment groups.

Reading Instruction Time

Across seven surveys, teachers in both the *PCI* and comparison group reported the number of minutes students received reading skills instruction in their classroom during a specified week. On average comparison teachers reported that students received 328 minutes per week and *PCI* teachers reported 270 minutes per week. In the *PCI* group, this number reflects both *PCI* instruction and instruction with other/supplemental reading materials.

Figure 8 shows the average minutes of reading instruction students received in their classroom over the course of the academic year surveyed. Comparison teachers reported more minutes of reading instruction than the *PCI* group during all surveyed weeks, except for March 2–6.

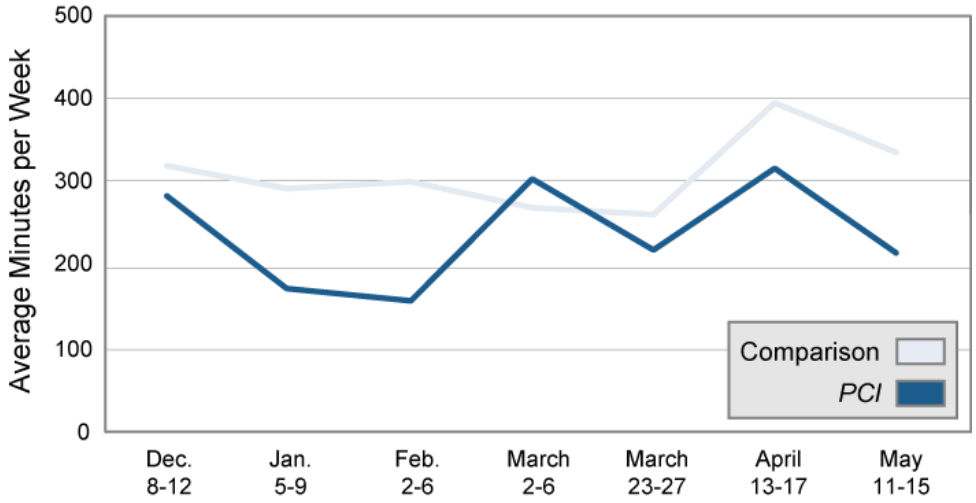

Figure 8. Average Weekly Minutes of Reading Instruction[18]

During Phase 1 we found a significant decrease in the use of *PCI* during the Florida Alternative Assessment (FAA) administration period. Therefore, we asked both assignment groups in Phase 2 whether they had stopped using or if they supplemented their current reading program in order to focus on FAA test preparation. Only one comparison teacher (8%) reported discontinuing the use of the current reading program in preparation for the FAA. However 5 out of 21 *PCI* teachers (24%) reported that they stopped *PCI* instruction to prepare for the FAA. Two other *PCI* teachers reported that they were not able to teach *PCI* to every student during this time and that they were supplementing instruction with other lessons.

Table 20. Have you stopped using or supplemented your current reading program (*PCI* for *PCI* group and existing reading program for comparison group) in order to focus on FAA test preparation?

	Yes, I was given explicit instructions by an administrator, head of department, etc.	Yes (other reason)	No, I am continuing to use only my current reading program in preparation for the FAA.	Other
Comparison (*n* = 13)	0 (0.0%)	1 (7.7%)	12 (92.3%)	0 (0.0%)
***PCI* (*n* = 21)**	2 (9.5%)	3 (14.3%)	12 (57.1%)	5 (23.8%)

[18] (N) counts varied across surveys.

Student Engagement

Researchers asked teachers in both groups to rate the average level of student engagement with their reading program. Teachers were instructed to consider students as fully engaged if they displayed consistent on-task behavior. Sixteen out of 20 *Level One* teachers (80%) and all of the *Level Two* teachers reported that their students were highly or very highly engaged with the program. In contrast, 8 out of 12 comparison teachers (67%) reported the same level of engagement.

Table 21. Level of Student Engagement

	Very high	High	Moderate	Low	Very low	I don't know
Comparison (*n* = 12)	3 (25.0%)	5 (41.7%)	3 (25.0%)	1 (8.3%)	0 (0.0%)	0 (0.0%)
***PCI Level One* (*n* =20)**	4 (20.0%)	12 (60.0%)	4 (20.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
***PCI Level Two* (*n* = 6)**	2 (33.3%)	4 (66.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)

PCI teachers were also asked to rate student level of engagement while participating in various aspects of the program. As shown in Figure 9 and Figure 10, a majority of teachers using *Level One* and *Level Two* reported that their students were highly or very highly engaged with the core steps of the lesson cycle.

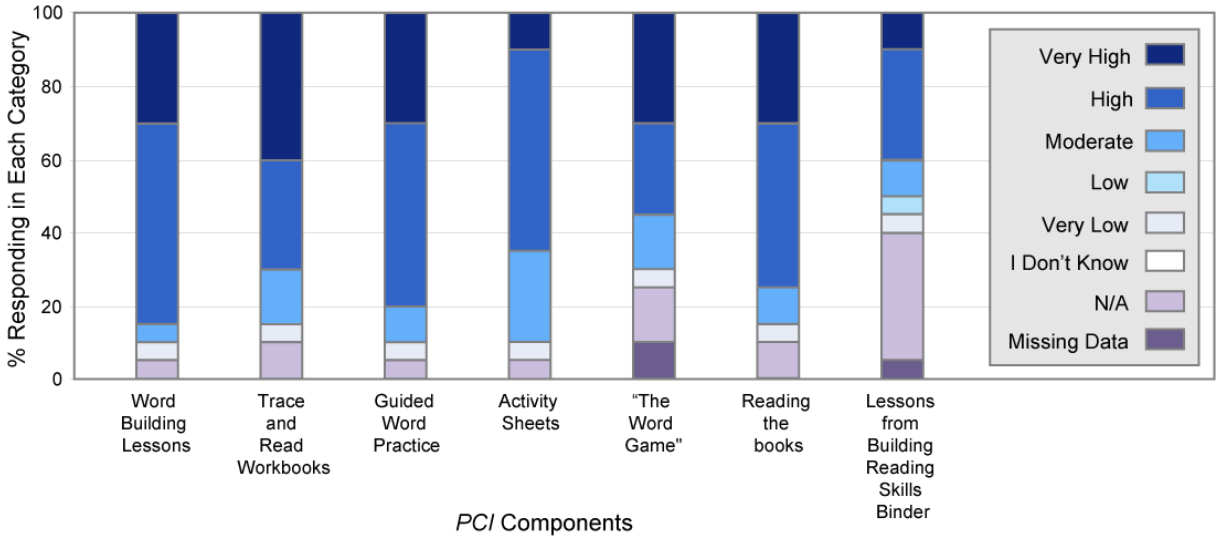

Figure 9. *Level One*: Levels of Student Engagement With *PCI* Components

Note. (n = 20)

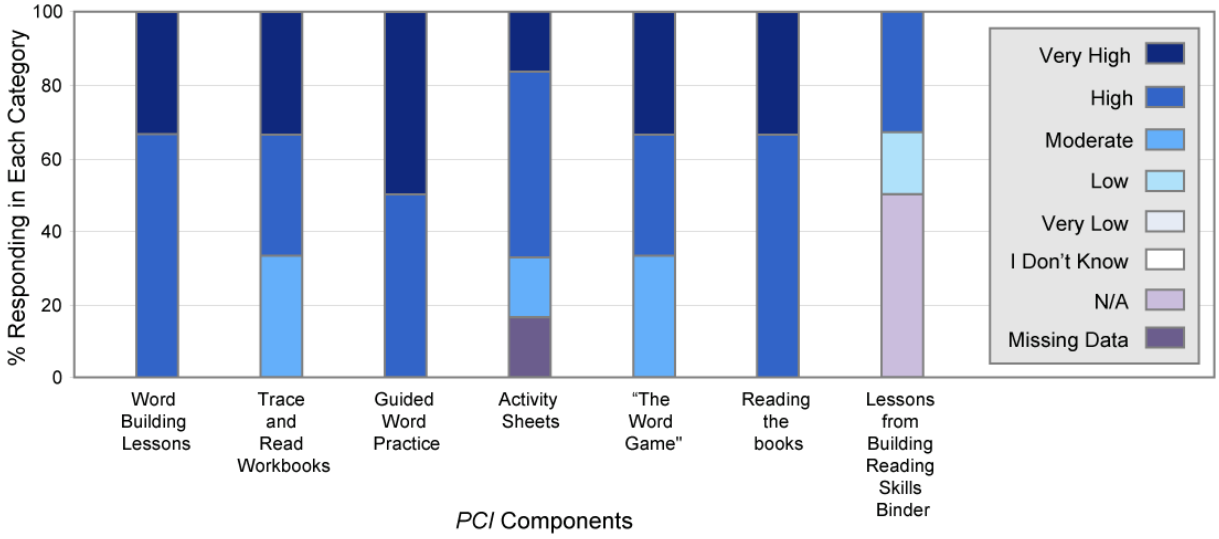

Figure 10. *Level Two*: Levels of Student Engagement With *PCI* Components

Note. (n = 6)

The word building lessons, guided word practice, and reading books were ranked very high in terms of student engagement, while many teachers were unable to rank student engagement in the lessons from the Building Reading Skills binder (possibly because they did not use this tool enough).

Summary

Overall, comparison teachers reported more minutes of reading instruction than *PCI* teachers reported; 24% of the *PCI* teachers reported that they discontinued use of the program during FAA administration. As in Phase 1, *PCI* teachers continued to report a high level of student engagement with *PCI*.

Student-Level Impact Results

In this section, we address the impact of the *PCI Reading Program* on student reading achievement. We present two kinds of analyses, quasi-experimental two-year impact and extra-experimental two-year impact (identified in the Analysis Plan section), in different subsections.

In both subsections, we examine three types of impacts:

- the student groups compared in the quasi-experimental two-year impact analysis
- the average impact of two years of *PCI* on the sight word assessment (which is the primary outcome measure)
- differential impacts of sight word pre-assessment and years of teaching Special Education across subgroups (for the quasi-experimental analysis); and differential impacts of sight word pre-assessment, phonological pre-assessment, and years of teaching Special Education across subgroups (extra-experimental analysis)

Additionally, in the extra-experimental subsection, we examine the difference in the impact of *PCI* on the sight word assessment based on different years of exposure to the program. We do not report the impact of *PCI* on phonological assessment for either the quasi-experiment or extra-experimental analysis. Phonological skills are introduced in *Level Two* of the program and, as described in the Implementation Results section, only six students progressed to *Level Two*. Therefore, there were not sufficient numbers of students to test the hypothesis that students' phonological skills would improve as they moved through the program.

Quasi-experimental Two-year Impact Analysis

Summary of Student Groups Compared in the Analysis

After limiting cases as described in the previous sections, our final sample for the quasi-experimental two-year impact analysis on sight word assessment consists of 74 students and 24 teachers. The *PCI* condition (illustrated as Group 1 in Figure 11) has 26 students and 11 teachers, and the comparison condition (illustrated as Group 3 and Group 5 in Figure 11) has 48 students and 13 teachers.

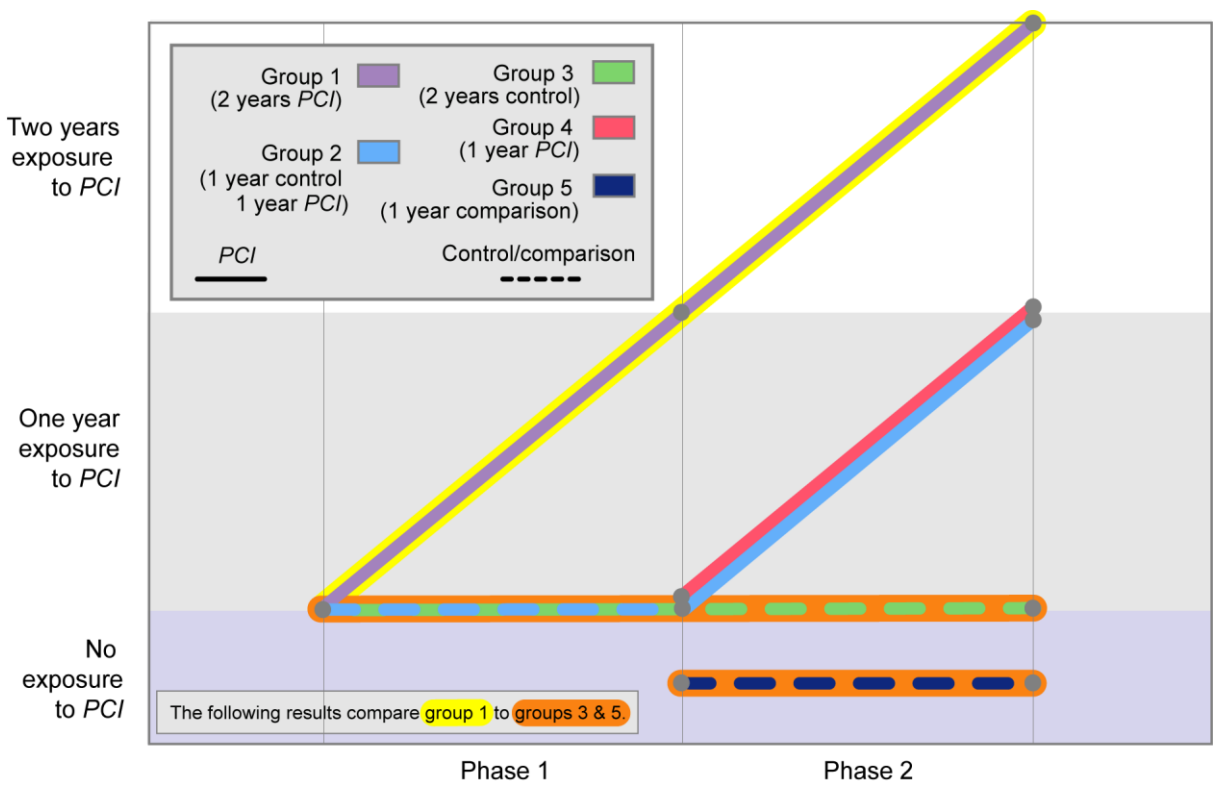

Figure 11. Student Groups Compared in Two-year Quasi-Experimental Analysis

Association of *PCI Reading Program* and Reading Achievement: Overall Score on the Sight Word Assessment

We first examine the effects of the *PCI Reading Program* on performance on the Sight Word Assessment. Table 22 provides a summary of the sample we used and outcomes on the Sight Word Assessment for students in *PCI* and comparison 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, classes, and teachers in that group. The last two columns provide the effect size, that is, the average difference in outcomes between the *PCI* and comparison groups in standard deviation units. Also provided is the *p* value, indicating the probability of arriving at a difference with an absolute value 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 original sample of students. (We removed one influential point from the analysis.[19]) The means, and therefore the effect size, are adjusted to take into account student sight word pretest scores, propensity scores, phonological pretest

[19] We used a critical value for Cook's Distance of 0.2.

scores as well as their grade levels; hence, these statistics are determined using matched samples with additional adjustment for imbalances on these covariates between the two groups.[20]

Table 22. Overview of Sample and Association of *PCI Reading Program* on Reading Achievement as Measured by the Sight Word Assessment

	Condition	Means	Standard deviations[a]	No. of students[c]	No. of teachers	Effect size	*p* value[b]	Percentile standing
Un-adjusted	**Comparison**	11.15	7.30	48	13	0.13	.64	5%
	PCI	12.08	6.30	26	11			
Adjusted	**Comparison**	11.15	7.30	48	13	0.89	.06	31%
	PCI	17.27	6.30	25	11			

[a] The standard deviations used to calculate the adjusted and unadjusted effect sizes are calculated from the posttest scores for students in the respective rows.

[b] The unadjusted effect size is Hedges' *g* with the *p* value adjusted for clustering of students in teachers. The adjusted effect size is the impact estimate from a PROC MIXED model that controls for clustering of students in teachers and that includes sight word pretest, propensity score, phonological pretest, and grade-level as covariates, divided by the estimate of the pooled standard deviation. The *p* value is for the impact estimate from that model.

[c] The sample sizes in this table show the numbers of cases retained after matching and removal of influential points.

Figure 12 provides a visual representation of the information in Table 22. 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 comparison and *PCI* groups. The pre- and posttest bars show that, on average, both the *PCI* and comparison groups grew in their Reading achievement (as measured by the sight word assessment) during the year.

The panel on the right is a visual display of results from the row labeled "Adjusted" in Table 22. It shows estimated performance on the posttest for the two groups based on a statistical equation that adjusts for students' pretest, propensity scores, and the other covariates. The overall effect size (in standard deviation units) is 0.89, which is equivalent to a gain of 31 percentile points for the median comparison group student if the student had received *PCI.* The fairly low *p* value for the *PCI* effect (.06) indicates we should have some confidence that the actual difference is different from zero. We added 80% confidence intervals to the tops of the bars in the figure. The non-overlap in these intervals further indicates that any difference we see is unlikely to be due to chance.

[20] In the analysis we first carry out propensity score matching to find comparison cases that are similar to program cases on a variety of dimensions. We then estimate the impact using the program group and the matched comparison cases through a statistical equation that adjusts for any further imbalance on several covariates as well as on the propensity score itself.

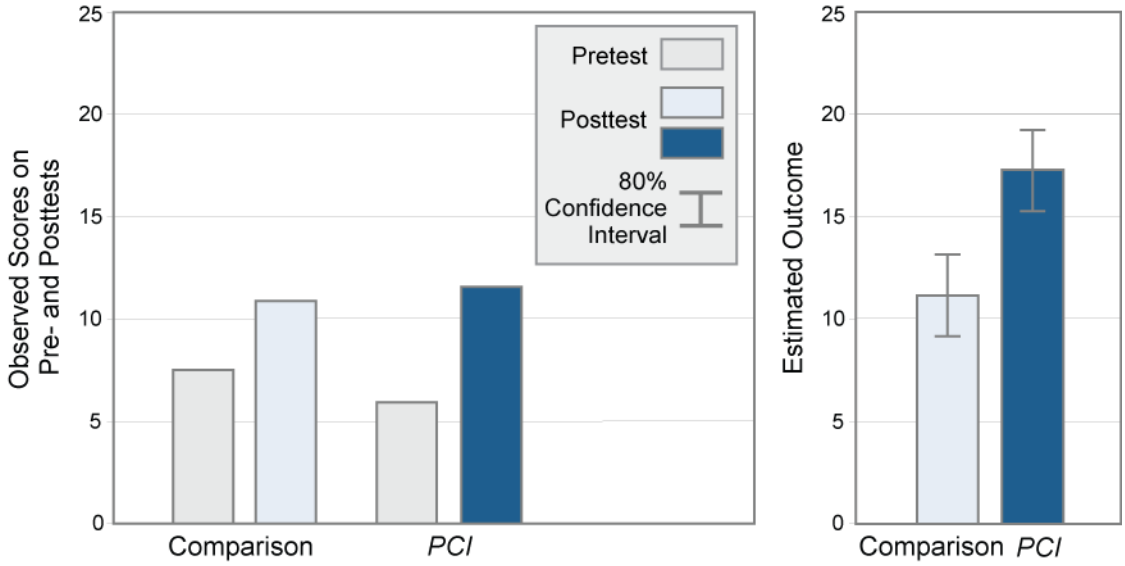

Figure 12. Relationship to Sight Word Recognition: Unadjusted Pre- and Posttest Means for Comparison and *PCI* (Left); Adjusted Means for Comparison and *PCI* (Right)

Moderating Variables

We now report the results of our analysis of the moderating effects of specific variables (Years of teaching special education and sight word pretest) on the impact of *PCI* on sight word recognition.[21] We had planned to look at the moderator effect of student phonological pretest scores, their disability status and ELL status, as well as their grade levels. However, there were too few students in the *PCI* or comparison group to make a comparison with sufficient statistical power.[22]

Including Sight Word Pre-Assessment as a Moderator

We first show whether the impact of *PCI* differs for students at various levels of prior achievement. At the bottom of the table we give results for technical review—these 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.

[21] The outcome of primary interest is performance on the sight word recognition scale. We therefore carry out moderator analyses for this outcome. We do not perform moderator analyses for the phonemic awareness outcomes.

[22] We did not perform the moderator analysis when the number of students in either the comparison or program group was less than eight.

Table 23 shows the estimated average difference between *PCI* and comparison in 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 23. The Moderating Effect of the Sight Word Pretest on the Impact of the *PCI Reading Program* on Sight Word Recognition

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
(1) Intercept	11.44	1.64	22	6.98	<.01
(2) Effect of *PCI* for a student with an average sight word pretest[b]	1.64	2.57	22	0.64	.53
(3) Change in outcome for a control student with a unit-increase in the sight word pretest[b]	0.89	0.19	40	4.20	<.01
(4) Change in outcome for a student with a unit-increase in the propensity score	-2.46	2.62	40	-0.94	.35
(5) Change in outcome for a student with a unit-increase in the phonological pretest[b]	0.42	0.20	40	2.09	.04
(6) Change in the effect of *PCI* with a unit-increase on the sight word pretest[b]	0.05	0.29	40	0.19	.85
Random effects[c]	**Estimate**	**Standard error**		***z* value**	***p* value**
Variance of teacher-level intercept	8.13	4.72		1.72	.04
Variance of teacher-level slope for pretest[d]	0.21	0.14		1.58	.06
Within-teacher variation	9.89	2.41		4.10	<.01

[a] We include as covariates in this analysis the pretest, the propensity score, and other covariates that were imbalanced between conditions. This includes grade level. We do not exhibit the effects associated with the different grade levels in the table.

[b] We center a given student's pretest score on the average pretest for the group to which that student is assigned (this is done for both the sight word and the phonological assessments). The pretests for the treatment and comparison groups were, for the most part, collected at different times; we chose not to model the component of the pretest that is completely confounded with time.

[c] Teachers were modeled as a random factor. Teacher-level variation in the intercept represents variation in teacher-level averages of student outcomes. Within-teacher variation represents the variability in student outcomes for each teacher.

[d] We set the average effect of pretest to vary randomly across teachers so that it reflects sampling variation, which is consistent with the way we model teacher-level intercepts. The *p* value for the interaction between treatment and the pretest therefore reflects uncertainty arising from the potential re-sampling of teachers and students.

The row in the table labeled "Effect of *PCI Reading Program* for a student with an average pretest" tells us whether *PCI* made a difference on Sight Word Assessment for a student who has a sight word pretest; that is, the same as the average sight word pretest in his/her respective condition. The estimate associated with *PCI* is 1.64. This shows a positive difference associated with *PCI*. The *p* value of .53 indicates that we can expect to see a difference, with an absolute value as large as or larger than the absolute value of the estimate, 53% of the time when there truly is no effect. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true impact is different from zero for the student who has a sight word pretest that is the same as the average pretest in his/her respective condition.

We also estimated the moderating effect of the pretest score[23] on the impact of *PCI* (row 6) to determine whether the intervention was differentially effective for students at different points along the pretest scale. The coefficient associated with the interaction of pretest with *PCI* is 0.05, which shows a very small increase in the *PCI* effect with each one-unit increase on the pretest. The *p* value of .85 indicates that we can expect to see a difference, with an absolute value as large as or larger than the absolute value of the estimate, 85% of the time when in fact there is zero impact—the small difference is easily the result of chance. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true differential impact is different from zero. In other words, the effect of *PCI* was the same for students, regardless of how well a student performed on the pretest.

As a visual representation of the results described in Table 23, we present a scatterplot in Figure 13, which shows student performance in reading, as measured by the Sight Word Assessment administered at the end of Phase 2, against their performance on sight word assessment in the fall 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, comparison students.

The two lines are the estimated values on the posttest for students in the *PCI* and comparison conditions. We observe an overall impact, but no significant difference in impact, across the prior score scale.[24]

[23] For the remainder of the quasi-experimental two-year impact analysis section, by the pretest we mean the condition-centered pretest (for both the sight word and phonological outcome); that is, we subtracted from each student's score the average pretest for the group (*PCI* or comparison group) to which that student belongs.

[24] The control group line applies to the 3rd grade comparison students who have an average propensity score and phonological pretest score. We examined the cases in the top left of the graph to determine whether they represented special cases who should potentially be removed from analysis. We did not find any obvious reason why they should be removed. As a consequence of randomization, such students present in both treatment and control groups. Their presence should not bias the estimate of mean impact.

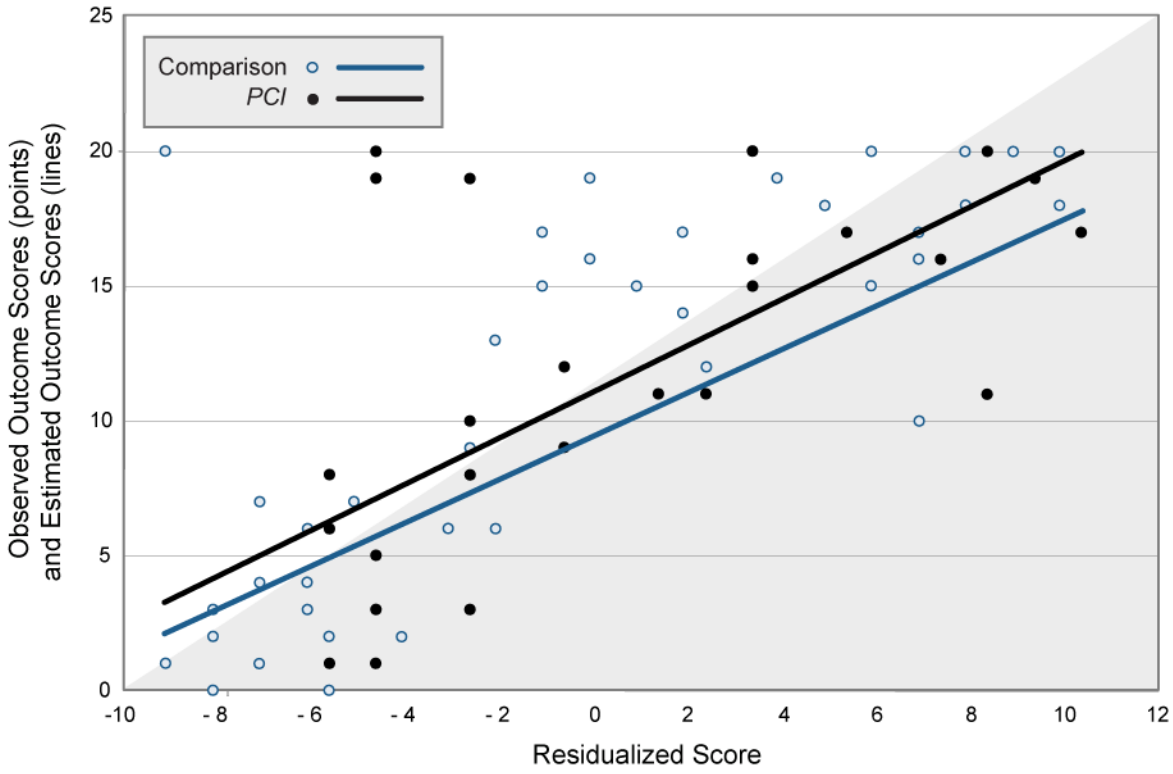

Figure 13. Comparison of Estimated and Actual Outcomes for *PCI* and Comparison Group Students (Sight Word Recognition)[25]

Including Years of Teaching Special Education as a Moderator

We also considered whether *PCI* is differentially effective for students who had relatively inexperienced Special Education teachers (three years or fewer) versus those with more experienced Special Education teachers (four or more years). Table 24 shows the results of our analysis of the moderating effect of years of teaching experience on students' performance on the Sight Word Assessment.

The row in the table labeled "Effect of *PCI* for the student whose teacher has 4+ years teaching experience" tells us whether *PCI* made a difference on sight word achievement for a student whose teacher has four or more years of Special Education teaching experience. The estimate associated with *PCI* Reading Program is 10.42. This shows a positive impact of *PCI*. The *p* value of .01 indicates that we have a high level of confidence that the true impact is different from zero for this subgroup.

The coefficient associated with the interaction of years of Special Education teaching experience (row 7) with *PCI* is -6.56, which shows a decrease in the *PCI* effect with three years or fewer of teaching experience. The *p* value of .04 indicates that we can expect to see a difference, with an absolute value as large as or larger than the absolute value of the estimate,

[25] We center a given student's pretest score on the average pretest for the group to which that student is assigned, thus the x-axis represents the condition-centered pretest score instead of the raw pretest score.

4% of the time when in fact there is zero differential impact. Using the criteria outlined earlier in the report, we conclude that we have strong confidence that the true differential impact is different from zero. In other words, the effect of *PCI* Reading Program was stronger for students whose teacher has more years of teaching experience in Special Education.[26]

[26] We emphasize the exploratory nature of this result due to the small number of teachers (11 *PCI* teachers and 12 comparison teachers).

Table 24. Moderating Effect of Years of Teaching Special Education on the Impact of *PCI* on Sight Word Recognition

Fixed effects[a]	Estimate	Standard error	DF	*t* value	*p* value
(1) Intercept	8.12	1.74	19	4.67	<.01
(2) Effect of *PCI* for a student whose teacher has 4+ years teaching experience	10.42	3.47	19	3.00	.01
(3) Change in outcome with a unit-increase in the sight word pretest score[b]	0.83	0.11	40	7.64	<.01
(4) Change in outcome with a unit-increase in the propensity score	-9.79	3.66	40	-2.67	.01
(5) Change in outcome with a unit-increase in the phonological pretest score[b]	-0.07	0.22	40	-0.33	.74
(6) Difference in control outcome between students whose teacher has <4 years teaching experience in Special Education and those whose teachers have 4+ years experience	3.91	1.67	19	2.34	.03
(7) Change in the effect of *PCI* between students whose teacher has <4 years teaching experience in Special Education and those whose teachers have 4+ years experience	-6.56	2.98	19	-2.20	.04

Random effects[c]	Estimate	Standard error	*z* value	*p* value
Teacher mean achievement	1.90	4.17	0.45	.32
Within-teacher variation	17.03	3.92	4.34	<.01

[a] We include as covariates in this analysis the pretest, propensity score, and other covariates that were imbalanced between conditions. This includes grade level. We do not exhibit the effects associated with the different grade levels in the table.

[b] We center a given student's pretest score on the average pretest for the group to which that student is assigned (this is done for both the sight word and the phonological assessments). The pretests for the treatment and comparison groups were, for the most part, collected at different times; we chose not to model the component of the pretest that is completely confounded with time.

[c] Teachers were modeled as a random factor. Teacher-level variation in the intercept represents variation in teacher-level averages of student outcomes. Within-teacher variation represents the variability in student outcomes for each teacher. We set the average effect of pretest to vary randomly across teachers so that it reflects sampling variation, which is consistent with the way we model teacher-level intercepts. In this analysis, the random slope is constrained to be zero by SAS and is not included in this table.

As a visual representation of the information presented in Table 24, the bar graph in Figure 14 shows the estimated difference between *PCI* and comparison conditions for students who had relatively inexperienced Special Education teachers (three years or fewer) versus those with more experienced Special Education teachers (four or more years). The bar graph includes

80% confidence intervals.[27] We focus here on the greater benefit of *PCI* when used by teachers with more than four years experience teaching Special Education. We emphasize that we have strong confidence that this trend is not the result of chance; however, as an exploratory result, it is deserving of continued follow-up through the next phases of the study.[28]

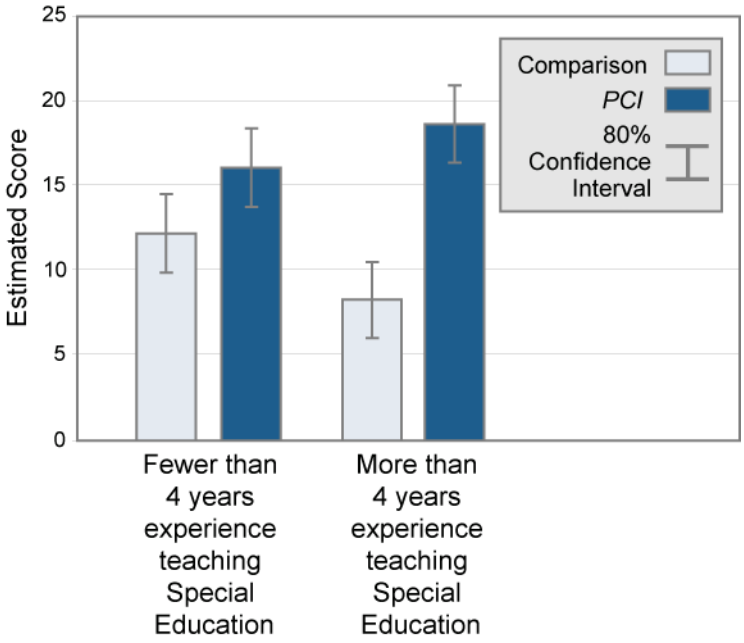

Figure 14. Moderating Effect of Years of Teaching Special Education on Sight Word Recognition for Two-year Impact

Mediator analysis

Researchers had planned to analyze whether the impact of *PCI* on sight word recognition was mediated by an intermediated impact of *PCI* on time spent teaching those skills. However, because we did not collect the minutes of instruction that each individual student received, we are not able to differentiate high-performing students from low-performing students within the

[27] The standard error for the interaction expresses uncertainty in the parameter that measures a difference in impact (i.e., the difference between levels of the moderator in the difference between program and control). The standard errors used for the confidence intervals in the graphs express uncertainty in the impact (i.e., the difference between program and control) at the given level of the moderator. The confidence intervals should be interpreted within but not across levels of the moderator.

[28] Note that all differences that are displayed apply to the sample of case in this analysis; however, to anchor the overall heights of the bars (rather than the differences among the bars, which are fixed by the statistical equation that we used), we use the estimate of average performance for students with average scores for the sight word pretest, the phonological pretest, and the propensity score. The overall heights are determined by average performance in 8th grade, which is the reference grade.

same class in terms of the amount of instruction they received. In other words, we cannot determine whether students with high scores received more instruction than students with low scores. Without student-level time data, we do not believe that the test would yield reliable results.

Extra-experimental Two-year Impact Analysis

We also obtain an estimate of the impact of two years of exposure to *PCI* on reading as measured through the sight word assessment using an extra-experimental method. This approach has an advantage that the quasi-experimental method does not: the estimate is not affected by selection bias. (Conversely, the quasi-experimental method has an advantage that the extra-experimental method does not: it is unaffected by bias due to the first-year program effect changing over time.) Combined, these two methods may provide convergent evidence of the impact of *PCI*—if the impact is in the same direction, then we have a stronger evidentiary base.

Stability in the first year of the program between Phase 1 and Phase 2 is an important requirement for the results of the extra-experimental analysis to be valid. *PCI* is an established program and, anecdotally, we did not see evidence of the intervention changing between Phase 1 and Phase 2 in a way that would lead to a difference between the two periods in a one-year impact. This leads us to consider the extra-experimental estimate as an especially important result that likely comes close to what an experimental estimate would have been, had we maintained a control group over both phases of the study. We examine whether the quasi- and extra-experimental analyses give convergent evidence of effectiveness.

Summary of Student Groups Compared in the Analysis

Based on the same criteria for limiting cases, our final sample for the extra-experimental two-year impact analysis on the sight word outcome consists of 40 students and 16 teachers. The *PCI* condition has 28 students and 11 teachers (illustrated as group 1 in Figure 16), and the comparison condition (illustrated as group 2 in Figure 15) has 12 students and five teachers.

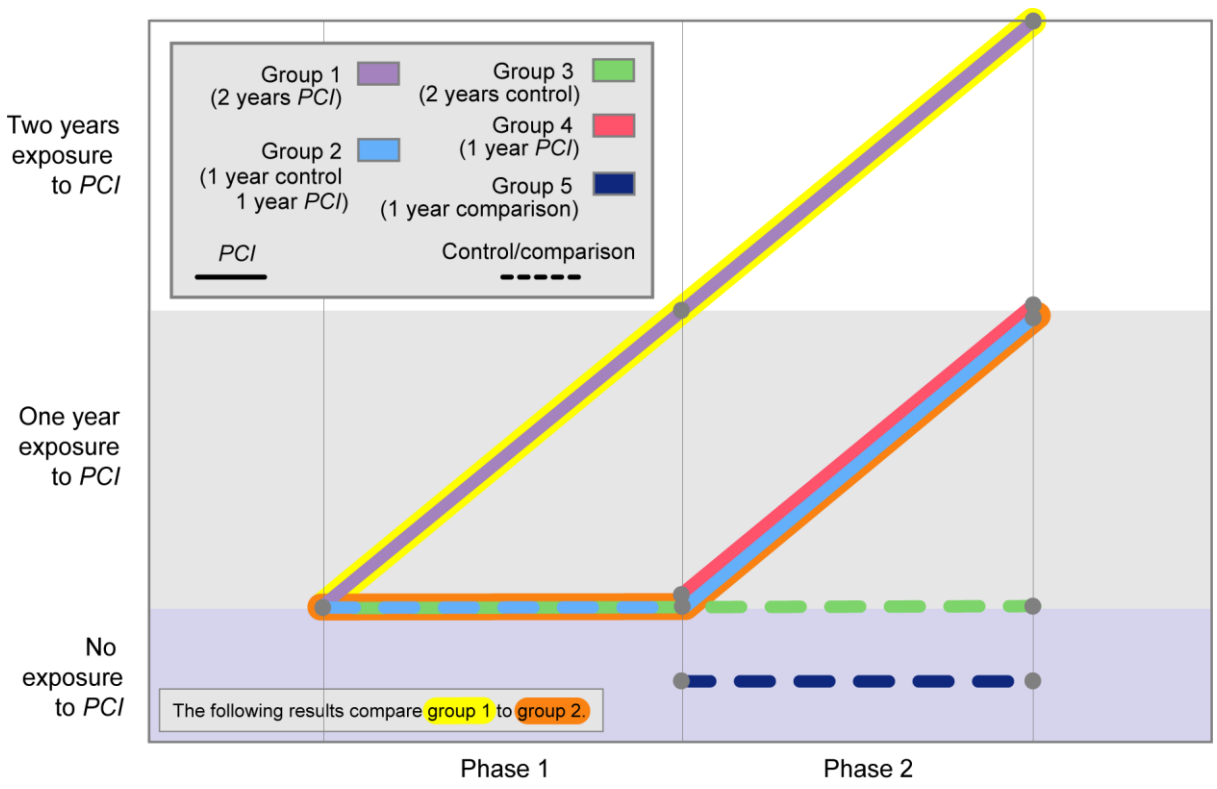

Figure 15. Student Groups Compared in Extra-Experimental Two-year Analysis

Impact of *PCI Reading Program* and Reading Achievement: Overall Score on the Sight Word Assessment

Table 25 provides a summary of the sample we used and the results for the sight word assessment. The group randomly assigned to receive *PCI* in Phase 1 (*PCI-1*) had an average posttest score 9.07 at the end of year 1. The control (*PCI*-2) had a regression-adjusted average score of 6.46[29]. The difference between these means is the unbiased estimate of the impact of *PCI* after one year of implementation. After two years, *PCI-1* had an average posttest score of 12.38. *PCI*-2 had a regression-adjusted average score of 9.18. The difference between these means is an estimate of the added effect of a second year of exposure to *PCI* (compared to only one year of exposure). The sum of the mean difference at the end of Phase 1 and the mean difference at the end of Phase 2 is the extra-experimental impact estimate. The value of this estimate is 5.81. The adjusted effect size is calculated by the dividing this estimate of the

[29] We have named the randomized group who receive *PCI* in Phase 1 as 'PCI-1' and the group randomized to control in Phase 1 and received *PCI* in Phase 2 as 'PCI-2'.

two-year impact by the estimate of the pooled standard deviation in posttest scores at the end of Phase 2. The impact estimate, and therefore the effect size, is adjusted to take into account the student pretest scores (obtained at the start of Phase 1); hence, these estimates are adjusted for any chance imbalances on the pretest between the two randomized groups. This effect size translates into a 34% difference in percentile standing. The *p* value of .02 corresponds to the two-year effect estimate and it indicates the probability of arriving at a difference with an absolute value as large as, or larger than, the absolute value of the one observed when there truly is no difference. The table also shows the sample sizes in the analytic sample[30] and the standard deviation in the outcomes for each group in each year[31, 32].

---

[30] To compute the two-year impact, we compare students who are receiving a second year of exposure to *PCI* to those who are receiving their first year exposure to *PCI*. We also compare outcomes for students who received exposure to *PCI* in the first year to those who did not. The two-year impact of *PCI* is measured in terms of the amount of exposure to the program that students are getting and not by the amount of exposure that teachers are receiving. In fact, the seven teachers who were replaced by new teachers in the experimental sample in the second year would be receiving a first year of exposure in Phase 2.

Classes of Special Education students remained largely intact from Phase 1 to Phase 2, which means that most of the students in a given class would be receiving *PCI* for the first or second year in Phase 2, depending on whether the class was originally assigned to *PCI* or control at the start of Phase 1. However, some students did switch classes, therefore, for example, there are cases of classes where most students are receiving *PCI* for the second year in Phase 2, but one student transfers in from a class that was not receiving *PCI* the previous year; therefore, that student would get his first exposure to *PCI* in Phase 2.

Given that the classes stay mostly intact, we examined whether the new teachers predominantly joined classes that consisted mostly of students who were receiving *PCI* for the first time or mostly of students who were receiving *PCI* for a second year. If the new teachers are systematically different from the teachers who left the study in a way that affects performance, and if they worked predominantly with students who are receiving *PCI* for the first time (or predominantly with students who are receiving *PCI* for a second year), then, as a result of this imbalance, the impact estimate may reflect the characteristics of the new teachers and therefore be biased as an estimate of the impact of the program. We verified that the new teachers were distributed fairly evenly among the two types of classes: those whose students were, for the most part, assigned to *PCI* the first year and those whose students were, for the most part, assigned to control the first year. We are therefore not concerned with this imbalance as a source of bias. (However, there is still a possibility of bias if some of the new teachers decided to work in the classes where *PCI* has already been in place and if those teachers have certain characteristics such as above-normal motivation.) Given the small number of new teachers, we cannot do a statistical check of whether the characteristics of the new teachers are evenly distributed between the two conditions.

We note that one teacher originally assigned to control insisted on not taking up *PCI* in Phase 2. She and her students are not included in the analysis of the extra-experimental impact. Her exclusion may introduce some bias but it is difficult to say in which direction the bias would go. Another teacher joined the study in Phase 2 and did not receive *PCI*. She received students who were in the control condition in year-1. This teacher also was not included in the analysis of the extra-experimental impact. The students of both of these teachers were included in the quasi-experimental analyses as possible comparison cases.

[31] The outcomes at the end of Phase 1 were computed by averaging the posttest for Phase 1 and the pretest for Phase 2. Where both scores were not available, we used the score that was available.

[32] The harmonic mean of the number of teachers is seven per condition. Using this value, and assuming an average of five students per class, as well as standard assumed values of specific parameters (an intraclass correlation of .15 and a pre-post correlation of .64), we compute that we can detect an effect size as large or

Table 25. Overview of Sample and Association of *PCI Reading Program* on Reading Achievement as Measured by the Sight Word Assessment: Extra-experimental Two-Year Impact Analysis

	Condition	Means	Standard deviations[a]	No. of students	No. of teachers	Effect size[b]	*p* value	Difference in percentile standing
Adjusted effect size	***PCI-1* (Phase 1)**	9.07	5.36	28	11	0.98	.02	34%
	***PCI-1* (Phase 2)**	12.38	6.30	26	11			
	***PCI-2* (Phase 1)**	6.46	4.02	12	5			
	***PCI-2* (Phase 2)**	9.18	4.77	11	5			

[a] The standard deviations used to calculate the adjusted effect sizes are calculated from the posttest scores for students in the respective rows.

[b] The adjusted effect size is based on the extra-experimental impact estimate obtained from SAS PROC MIXED where we model clustering of students in teachers and that includes the sight word pretest as a covariate divided by the estimate of the pooled standard deviation in the outcome scores at the end of Phase 2. The *p* value is for the impact estimate from that model.

Figure 16 provides a visual representation of the analysis results. The bar graph displays the average reading outcomes as measured by the sight word assessment. Specifically, the bars on the left show the predicted averages (regression-adjusted means based on a statistical equation that adjusts for students' pretest, propensity scores, and other background covariates) of the posttest scores for the control and *PCI* groups at the end of Phase 1. The difference in the height of the bars represents the one-year impact of *PCI*. The bars on the right show the predicted averages of the posttest scores for the control and *PCI* groups after two years, assuming the control group received no *PCI* instruction in Phase 2. The difference in the height of the bars represents the two-year impact of *PCI*. As suggested by the bar graph, on average the *PCI* group made consistent progress in their Reading achievement, as measured by the sight word assessment, over the two years. We added 80% confidence intervals to the tops of the bars in the figure. The non-overlap in these intervals further indicates that any difference we see between the control and *PCI* groups is unlikely to be due to chance.

---

larger than .80 standard deviation with 80% power and assuming a type-1 error rate of .05. The effect observed is large and, as a consequence, we are able to detect it with a relatively small sample of teachers.

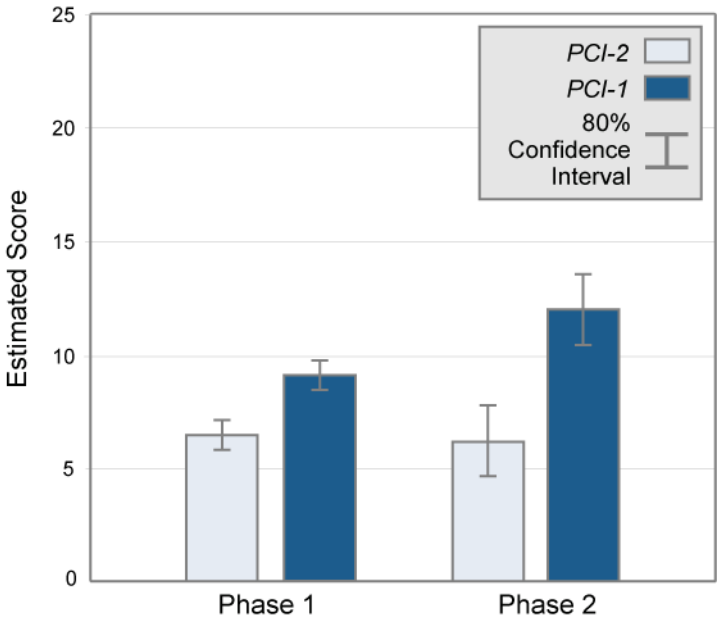

Figure 16. Impact on Sight Word Recognition: Year 1 Impact (Left); Year 2 Impact (Right)

Testing the Effect of an Additional Year of Exposure

The extra-experimental impact estimate consists of the average difference in performance between the two randomized groups at the end of Phase 1 plus the average difference in performance between the two groups at the end of Phase 2. The former of these is an unbiased estimate of the impact of one year of exposure to *PCI* compared to no exposure to *PCI*. The latter is an unbiased[33] estimate of the effect of two years of exposure to *PCI* compared to one year of exposure. The two-component estimate allows us to address two further questions.

Does the benefit of PCI continue beyond the first year? If so, then we expect the students with two years' exposure to *PCI* to outperform the students with one year exposure to *PCI*. In other words, we can subtract-off the one-year experimental impact estimate from the two-year extra-experimental impact estimate and check whether the remaining term is statistically significant. We observe that the difference between the groups at the end of Phase 2 is 3.20 scale score units (p = .06[34]), which gives us some confidence that the average difference in sight word performance between the groups at the end of Phase 2 is statistically significant.

Does the difference between the two groups in average sight word performance change from Phase 1 to Phase 2? The answer to this question allows us to tell whether the effect of *PCI* is accruing at a decelerating rate. The estimated difference between the Phase 1 and Phase 2 differences is .60 scale score units (p = .65), which gives us no confidence of a significant

[33] Lack of bias in the end-of-year estimates of the difference between groups depends partially on the potential biasing effects of attrition. We provide a brief analysis of attrition in a later section.

[34] The average difference in sight word performance at the end of Phase 2 is larger than at the end of Phase 1, but the p value is slightly larger. Greater uncertainty in the estimate at the end of Phase 2 (compared to Phase 1) is due to increasing variance in performance among students over time within each condition.

change in the differences between the two groups. At the end of each phase, the group with more *PCI* exposure is outperforming the group with less *PCI* exposure, and to the same extent. This suggests that, over the two years, the accrual of the *PCI* effect is positive and linear—it is not diminishing.

Moderating Variables

We now report the results of our analysis of the moderating effects of specific variables (Years of teaching special education, phonological pretest, and sight word pretest) on the impact of *PCI* on sight word recognition.[35] We had planned to look at the moderator effect of student autistic status and ELL status as well as their grade levels. However, there were not sufficient numbers of students in the *PCI* or comparison group to make a powerful statistical comparison.[36]

The moderator analyses for the extra-experimental impact estimates are based on extensions of the statistical equations used to calculate the average impacts. We report only the effect estimate of interest.

Including Sight Word Pre-Assessment as a Moderator

We used sight word performance collected at the start of Phase 1 to measure whether there was a difference in impact depending on a student's starting performance. The change in impact associated with a one-unit increase in the sight word pretest was -.33. The *p* value of .57 indicates that we can expect to see a difference, with an absolute value as large as or larger than the absolute value of the estimate, 57% of the time when there truly is no differential effect. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true impact is different from zero when controlling for the moderator effect of the sight word pretest.

Including Phonological Pre-Assessment as a Moderator

We used phonological performance collected at the start of Phase 1 to measure whether there was a difference in impact depending on a student's starting performance. The change in impact associated with a one-unit increase in the sight word pretest was .98. The *p* value of .30 indicates that we can expect to see a difference, with an absolute value as large as or larger than the absolute value of the estimate, 30% of the time when there truly is no differential effect. Using the criteria outlined earlier in the report, we conclude that we have no confidence that the true impact is different from zero when controlling for the moderator effect of the phonological pretest.

Including Years of Teaching Special Education as a Moderator

We also considered whether *PCI* is differentially effective for students who had relatively inexperienced Special Education teachers (three years or fewer) versus those with more experienced Special Education teachers (four or more years). The change of impact associated

[35] The outcome of primary interest is performance on the sight word recognition scale. We therefore carry out moderator analyses for this outcome. We do not perform moderator analyses for the phonemic awareness outcomes.

[36] We did not perform the moderator analysis when the number of cases in either the comparison or program group was less than eight students.

with having a teacher with three of fewer years of teaching experience is -3.72, which shows a decrease in the *PCI* effect. The *p* value of .49 indicates that we can expect to see a difference, with an absolute value as large as or larger than the absolute value of the estimate, 49% of the time when in fact there is zero differential impact. Using the criteria outlined earlier in the report, we conclude that we have little confidence that the true differential impact is different from zero.

Summary of Student-Level Impact Results

As noted earlier in the report, we obtain a stronger evidence base when we can address several questions using complementary methods, where each method has strengths that another method may not have. If the results are consistent for the different methods we can have greater confidence in them than if we use a single method, or if we use more than one approach and get inconsistent results. The following tables summarize the student impact results of Phase 1 and Phase 2 of this study. Table 26 summarizes the main impact results and addresses bias due to selection, changing treatment, and attrition. Table 27 summarizes the moderating effects of certain variables[37] (Jaciw, Zheng 2010).

Table 26. Summary of Main Impacts

Outcome	Impact	Method	Estimate[a]	Bias due to selection	Bias due to a changing treatment	Bias due to attrition
Sight word	1-year	Experimental	3.17 ($p < .05$)	Ruled out	Ruled out	Unlikely
Sight word	2-year	Quasi-experimental	6.12 ($p = .06$)	Passes internal checks, but possible	Ruled out	Passes internal checks, but possible
Sight word	2-year	Extra-experimental	5.81 ($p = .02$)	Ruled out	Unlikely	Passes internal checks, but possible

[a] In units of the outcome scale

[37] In this study, the two-year impact estimates from both methods are consistent, which gives us convergent validity and greater confidence in the result. We compared the two methods and discussed their strengths and limitations for estimating long-term program impacts (when the control group joins treatment in the short-term) at the 2010 conference of the Society for Research on Educational Effectiveness (SREE).

Table 27. Summary of Moderating Effects of Specific Variables

	Sight word pretest	Phonological pretest	Fewer than 4 years teaching Special Ed.	Disability	ELL	Grade level
Experimental (1-year)	-0.15 (*p* = .25)	0.35 (*p* = .11)	N/A[1]	Did not conduct	Did not conduct	0.27 (*p* = .52)
Quasi-experimental (2-year)	0.05 (*p* = .85)	N/A[a]	-6.56 (*p* = .04)	N/A[b]	N/A[b]	N/A[b]
Extra-experimental (2-year)	-0.33 (*p* = .57)	0.98 (*p* = .30)	-3.72 (*p* = .49)	N/A[b]	N/A[b]	N/A[b]

[a] Analysis not run because of imbalance between *PCI* and comparison groups.

[b] Insufficient power to run this analysis.

Discussion

Overview

This report contains the findings from Phase 2 of a five-year longitudinal study on the efficacy of the *PCI Reading Program* as implemented in two Florida school districts. The study both extends the original experimental design and analysis and introduces a comprehensive matched quasi-experimental design to investigate whether students whose teachers used *PCI* achieved higher sight word and phonological assessment scores than students whose teachers did not have the program. Researchers investigated whether *PCI* had a different effect on sight word recognition for specific subgroups of students: those who scored low on the sight word and phonological pre-assessments and those whose teachers had more experience teaching Special Education. Our sample was composed of students with supported level disabilities in grades 3–8 and their teachers from Brevard Public Schools and Miami-Dade County Public Schools. Our outcome measures were a sight word assessment, developed by researchers in collaboration with independent consultants, and a modified version of DIBELS Initial Sound Fluency assessment.

As in Phase 1, we continued to approach this experiment as an efficacy trial. 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 sight word pre- and posttest consisted of a sample of words taken from *Level One* of the *PCI* program itself. It was not a general test of reading but rather one that was closely aligned to the program. Phonological skills are introduced in *Level Two* of the program and are important skills to have for *Level Three.* Therefore, we wanted to track this outcome as students progress through the program.

Due to the study design and the criteria set for teacher participation, students in Phase 2 had different levels of exposure to the program (see Table 6 and Figure 1 for reference of student groups). This allowed researchers to conduct two different analyses to estimate the two-year impact of *PCI*: 1) using a quasi-experimental approach comparing assessment scores of students who had received *PCI* instruction for two years to assessment scores of students who had not been exposed to *PCI* and 2) using an extra-experimental approach comparing assessment scores of *PCI* students who were part of the randomized *PCI* group in Phase 1 to scores of Phase 1 control students who used *PCI* in Phase 2.

Student Impact Results

In both approaches to estimating the two-year impact (quasi-experimental and extra-experimental), we found that students in *PCI* classrooms performed significantly higher on the sight word outcome assessment (adjusted effect sizes of 0.89 and 0.98, respectively) with small *p* values (.06 and .02, respectively). The difference found in the quasi-experiment is equivalent to a difference of 31 percentile points and, in the extra-experimental, approach is equivalent to a difference of 34 percentile points. We also found that students continue to grow in sight word recognition with a second year of exposure to the program and that the effect of *PCI* is larger after two years than it is after one year. We did not report the impact of *PCI* on the phonological outcome assessment because there were very few students on *Level Two* of the program, where these skills are introduced.

In examining moderator effects using the quasi-experimental approach, we found the sight word pre-assessment to not be a significant moderator of the impact on sight word post-assessment. However, we have strong confidence that students whose teachers have more than four years of Special Education teaching experience benefit more from *PCI* than students with teachers who have fewer than four years of Special Education teaching experience. In examining moderator effects using the extra-experimental approach, we found no significant moderating effects. However, these analyses may be underpowered, given the small sample sizes in the program and comparison groups, and deserve additional exploration. Due to sample size issues, we were unable to examine the effects of

other moderating variables (student who were autistic, English language learners, and students in lower or higher grades).

Implementation Results

Overall implementation conditions across both groups were comparable and generally good for implementing *PCI*. As in Phase 1, during trainings, classroom observations, and on surveys, *PCI* teachers continued to show enthusiasm for and satisfaction with the program; they continued to report a high level of student engagement with *PCI*. Almost all teachers said they would continue using the program after the completion of the study. However, a majority of the *PCI* teachers supplemented reading instruction with other materials. Additionally, nearly half of the students who began on *Level One* at the beginning of the year remained on words 1-20 by the end of the year. While teachers were satisfied and students were engaged with the program, the lack of student progress through the program, and use of additional reading materials, may suggest that teachers were not able to complete the ideal minutes of instruction per student, as intended by the publisher. Because we did not measure time spent on the program on a per-student basis, we are unable to test this assumption. However, many teachers did report that the primary difficulty in implementing *PCI* was finding the time for individualized instruction. Principals and teachers in both groups expressed a concern about the lack of curricula available for supported level classrooms. *PCI* teachers were encouraged that the program helps fulfill that need.

Conclusion

This study included a complex design in which we conducted multiple analyses estimating the efficacy of the *PCI Reading Program.* The considerably large two-year impact on the sight word assessment, found using both the quasi- and extra-experimental approaches, gives us strong evidence that *PCI* students are learning to read the words in the *PCI Reading Program,* especially as their experience with the program increases. However, we also found that a majority of the students were not progressing through the program at the rate expected by the program developers. Teachers reported difficulty finding time to implement the individualized components of *PCI* and meeting the minimum time requirements. Because we were unable to measure individual student usage of the program, we do not know whether students who received more *PCI* instruction progressed farther through the program or benefitted more from *PCI*. We also lacked the sample size and background information about the students' ability level and capacity to learn that would be necessary to investigate whether *PCI* had a differential impact for varying classifications of students with supported level disabilities.

Outcomes for these districts should continue to be tracked as teachers continue to use and students progress through the program. Additionally, future studies would benefit from more general measures of reading as well as more detailed records of individual student characteristics, including program usage at the individual student level, allowing the study to explore the optimal implementation modes and conditions for success.

References

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

Bell, S. H., & Bradley, M. C. (2008, March). *Calculating long-run impacts in RCTs that release the control group into the intervention prior to the end of follow-up.* Paper presented in a paper discussion at the Annual Research Conference of the Society for Research on Educational Effectiveness, Crystal City, VA.

Florida Department of Education (2008, May). Profiles of Florida School Districts 2006-2007. Retrieved on January 16, 2009 from http://www.fldoe.org/eias/eiaspubs/pdf/ssdata07.pdf

Jaciw, A., Zheng, X., Ma, B., & Zhao, Q. (2010, March). *Estimating Long-Term Program Impacts When the Control Group Joins Treatment in the Shortterm: A Theoretical and Empirical Study of the Tradeoffs between Extra- and Quasi-Experimental Estimates.* Poster session presented at the annual research conference of the Society for Research on Educational Effectiveness, Washington, DC.

Shadish, W. R., Luellen, J. K., & Clark, M. H., (2006) Propensity scores and quasi-experiments: *A testimony to the practical side of Lee Sechrest.* In R.R. Bootzin & P.E., McKnight (Eds.), Strengthening research methodology: Psychological measurement and evaluation (pp. 143-157). Washington DC: American Psychological Association.

The Center for Public Education (2005, October 4). Teacher quality and student achievement: Key lessons learned. Retrieved December 1, 2008, from http://www.centerforpubliceducation.org/site/c.kjJXJ5MPIwE/b.1510943/k.9EF1/Teacher_quality_and_student_achievement_ Key_lessons_learned.htm

Toby, M., Ma, B., Jaciw, A., & Cabalo, J. (2008, October). *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.* (Empirical Education Rep. No. EEI_*PCI*-07100-FR1-Y1.O1). Palo Alto, CA: Empirical Education Inc.

U.S. Census Bureau (2008, July 25). 2006 Estimated Census Population for Brevard County, Florida. Retrieved on January 16, 2009 from http://quickfacts.census.gov/qfd/states/12/12009.html

U.S. Census Bureau (2008, July 25). 2006 Estimated Census Population for Miami-Dade County, Florida. Retrieved on January 16, 2009 from http://quickfacts.census.gov/qfd/states/12/12086.html

Appendix

(*n*) counts are not presented in the tables because the percentages represent an average across three surveys. Results may equal less than 100% if teachers did not respond to a specific category within the question and their data was therefore considered missing.

Table 28. *Level One*: Which steps do you usually complete during each lesson cycle?

	Always	Sometimes	Never	NA
Step 1a. Learn a Word: Word Building Lesson	84.4%	13.1%	0.0%	2.5%
Step 1b. Learn a Word: Trace and Read Workbook	76.7%	21.9%	0.0%	2.5%
Step 1c. Learn a Word: Guided Word Practice	76.3%	21.3%	0.0%	2.5%
Step 1d. Learn a Word: Activity Sheets	59.4%	35.%	0.0%	2.5%
Step 2. Repeat steps 1a-1d to learn four new words	48.1%	35.6%	8.1%	5.6%
Step 3. Review the words with "The Word Game"	24.4%	33.1%	22.5%	14.4%
Step 4. Assess Word Retention	53.1%	27.5%	5.6%	5.6%
Step 5. Read a Book	58.8%	26.9%	3.1%	11.3%

Table 29. *Level Two*: Which steps do you usually complete during each lesson cycle?

	Always	Sometimes	Never	NA
Step 1a. Learn a Word: Word Building Lesson	100.0%	0.0%	0.0%	0.0%
Step 1b. Learn a Word: Trace and Read Workbook	100.0%	0.0%	0.0%	0.0%
Step 1c. Learn a Word: Guided Word Practice	87.5%	12.5%	0.0%	0.0%
Step 1d. Learn a Word: Activity Sheets	100.0%	0.0%	0.0%	0.0%
Step 2. Repeat steps 1a-1d to learn four new words	77.5%	22.5%	0.0%	0.0%
Step 3. Review the words with "The Word Game"	12.5%	77.5%	10.0%	0.0%
Step 4. Assess Word Retention	100.0%	0.0%	0.0%	0.0%
Step 5. Read a Book	87.5%	12.5%	0.0%	0.0%

Table 30. *Level One*: 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	81.2%	13.8%	0.0%	2.5%
Step 1b. Learn a Word: Trace and Read Workbook	64.4%	24.4%	28.8%	2.5%
Step 1c. Learn a Word: Guided Word Practice	86.9%	8.1%	0.0%	2.5%
Step 1d. Learn a Word: Activity Sheets	55.6%	22.5%	35.6%	5.0%
Step 2. Repeat steps 1a-1d to learn four new words	67.5%	19.4%	2.5%	18.8%
Step 3. Review the words with "The Word Game"	29.4%	15.6%	3.1%	33.8%
Step 4. Assess Word Retention	78.1%	2.5%	0.0%	13.8%
Step 5. Read a Book	78.1%	16.3%	8.1%	8.1%

Note. Because teachers could select more than one category, totals may exceed 100%.

Table 31. *Level Two*: 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	90.0%	10.0%	0.0%	0.0%
Step 1b. Learn a Word: Trace and Read Workbook	30.0%	20.0%	35.0%	0.0%
Step 1c. Learn a Word: Guided Word Practice	77.5%	10.0%	0.0%	0.0%
Step 1d. Learn a Word: Activity Sheets	30.0%	20.0%	35.0%	0.0%
Step 2. Repeat steps 1a-1d to learn four new words	100.0%	10.0%	0.0%	0.0%
Step 3. Review the words with "The Word Game"	35.0%	30.0%	0.0%	10.0%
Step 4. Assess Word Retention	100.0%	0.0%	0.0%	0.0%
Step 5. Read a Book	100.0%	0.0%	0.0%	0.0%

Note. Because teachers could select more than one category, totals may exceed 100%.

Table 32. *Level One*: Who is this step usually taught by?

	Teacher	Other adult	NA
Step 1a. Learn a Word: Word Building Lesson	75.6%	32.5%	2.5%
Step 1b. Learn a Word: Trace and Read Workbook	59.4%	52.5%	2.5%
Step 1c. Learn a Word: Guided Word Practice	78.1%	35.6%	2.5%
Step 1d. Learn a Word: Activity Sheets	67.5%	46.9%	5.0%
Step 2. Repeat steps 1a-1d to learn four new words	65.0%	27.5%	18.8%
Step 3. Review the words with "The Word Game"	46.9%	32.5%	30.6%
Step 4. Assess Word Retention	61.9%	21.9%	16.3%
Step 5. Read a Book	66.9%	49.4%	8.1%

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

Table 33. *Level Two*: Who is this step usually taught by?

	Teacher	Other adult	NA
Step 1a. Learn a Word: Word Building Lesson	20.0%	77.5%	0.0%
Step 1b. Learn a Word: Trace and Read Workbook	10.0%	40.0%	12.5%
Step 1c. Learn a Word: Guided Word Practice	20.0%	77.5%	0.0%
Step 1d. Learn a Word: Activity Sheets	20.0%	20.0%	35.0%
Step 2. Repeat steps 1a-1d to learn four new words	30.0%	55.0%	0.0%
Step 3. Review the words with "The Word Game"	20.0%	45.0%	10.0%
Step 4. Assess Word Retention	67.5%	20.0%	0.0%
Step 5. Read a Book	57.5%	77.5%	0.0%

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