Empirical Education

RESEARCH REPORT

PCI Reading Program:

The Final Report of a Three Year Experimental Study in Brevard Public Schools and Miami-Dade County Public Schools

Megan Toby
Andrew Jaciw
Boya Ma
Akiko Lipton

Empirical Education Inc.

July 2011

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

Acknowledgements

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

About Empirical Education Inc.

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

Executive Summary

Introduction. PCI Education contracted with Empirical Education Inc. to conduct a three-year longitudinal study to determine the comparative effectiveness of the *PCI Reading Program (PCI)* for students with severe disabilities as implemented in Florida's Brevard Public Schools and Miami-Dade County Public Schools. The primary question addressed by the study is whether students whose teachers have received the *PCI Reading Program* improve their reading skills as indicated by higher Sight word and Phonological assessment scores than students whose teachers have not had exposure to the program. The Sight word pre- and posttest consisted of a sample of words taken from the *PCI* program itself. It was not a general test of reading but rather one that was closely aligned to the program.

The study began in Phase 1 (2007-2008 school year) as a randomized control trial. In Phase 2, students in experimental control classes joined the *PCI* group, and a new set of comparisons students was recruited. 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 the comparison students who had not been exposed to *PCI* and 2) using an extra-experimental approach based on the differences in performance at the end of Phases 1 and 2 for students who were members of the originally randomized classes. 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. In Phase 3, the small remaining sample size of students prevented researchers from being able to determine an impact on the Sight word or Phonological scores as was done in Phases 1 and 2 of the study. Instead, researchers investigated associations between student progress and years of participation in the program, while continuing to track the implementation of the program.

Findings. In Phase 1, the experiment was able to detect an impact on sight word learning equivalent to a 21 percentile point difference between students in the *PCI* program and those in the comparison group. Both the unadjusted analysis and the analysis where we adjusted for the effects of covariates show high effect sizes of .55 and .59 standard deviation units, respectively, with small *p* values. Since Level One of the program does not teach phonological skills, no phonological posttest was administered during Phase 1 and researchers did not examine an effect on phonological skills until Phase 2 of the study.

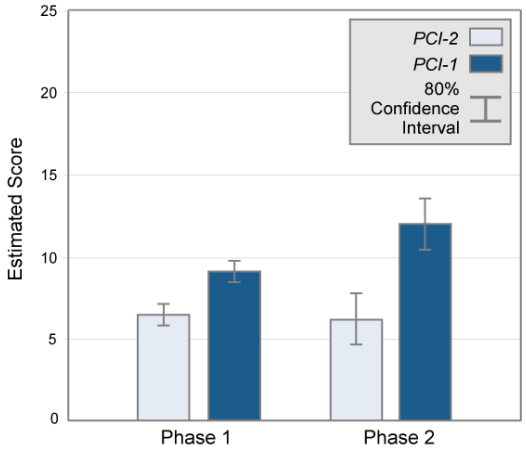

Figure 1. Impact on Sight Word Recognition Using Extra-Experimental Approach: Year 1 Impact (Left); Year 2 Impact (Right)

Phase 2 confirmed the significant positive findings from Phase 1. In both the quasi-experimental and extra-experimental approaches to estimating the two-year impact of *PCI*, researchers 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, researchers found that students continue to improve their sight word recognition 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 phonological skills because very few students progressed to *Level Two*- the program level in which phonological skills are introduced. Additionally,

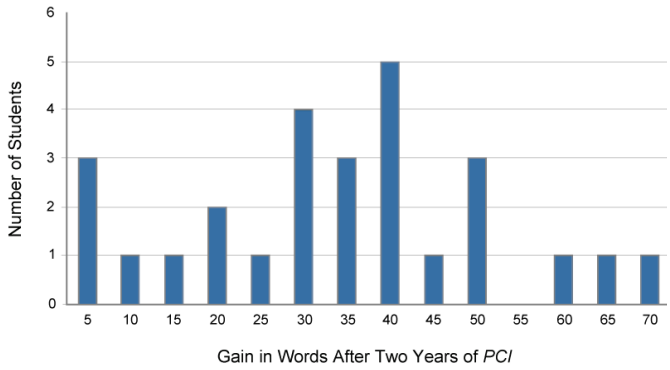

Figure 2. Word Gain After Two Years

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 displays the main findings from Phases 1 and 2.

Due to the small number of students remaining in the *PCI* condition, researchers were not able to conduct analyses of the three year impact at the end of Phase 3. However, analysts were able to measure the number of words students actually mastered through the course of the program (end word minus start word). Figure 2 displays the number of words students gained over two years of exposure to the *PCI Reading Program*. We found that the median students gained 33 words over the two-year span. With one year of exposure, the median student gains 14.5 words, as opposed to the 140 words as initially predicted by the publisher. The rate at which students learn words increases to 16.5 after a second year of exposure to the program.

Overall Teacher Impressions. Phase 3 of the study corroborated most of what researchers had learned, with regard to program satisfaction, from Phases 1 and 2. Teachers continued to feel well prepared for teaching *PCI* after the initial training, and general conditions for implementing the program—including access to all needed materials—were very good. Throughout all three phases of the study teachers consistently reported high levels of satisfaction with the program, intent to continue teaching *PCI* at the conclusion of the research study, and an endorsement of the program to other reading teachers of similar populations. In addition, teachers reported high levels of student engagement and perceived enjoyment with *PCI*. Teachers across all three phases continued to supplement *PCI* instruction with a variety of other materials. Teachers also consistently commented on the struggle to find time for the individualized instruction required for ideal implementation.

Design and Analysis. The study began in Phase 1 as a randomized control trial with teacher-level randomization. Phase 2 used a matched quasi-experimental design, comparing assessment scores of students who had received exposure to *PCI* for two years to students who had received no exposure to *PCI*. We also used an extra-experimental design to estimate the two-year impact of *PCI*, which compared assessment scores of *PCI* students who were part of the randomized *PCI* group in Phase 1 to scores of 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. The impact estimates were adjusted for any chance imbalances on relevant students and teachers characteristics between the two groups. The two complimentary methods produced consistent impact estimates which provided us with convergent validity and greater confidence in our results.

Information on student and teacher background characteristics as well as program implementation was gathered through online surveys, observations, and teacher interviews. Data regarding student start and end words were obtained through teacher surveys.

Conclusion. This study provides evidence of the efficacy of the *PCI Reading Program*. The significantly large impact found in the first two phases of the research study and high levels of teacher satisfaction with the program provides useful information for school districts looking for a reading program for severely disabled students. While student progress through the program is slower than initially expected by the publisher, students learn and retain the sight words explicitly taught by the program and benefit from continued years of exposure.

***PCI Reading Program*:**

The Final Report of a Three Year Experimental Study in Brevard Public Schools and Miami-Dade County Public Schools

Table of Contents

Introduction

PCI Education contracted with Empirical Education Inc. to conduct a three-year longitudinal study to determine the comparative effectiveness of the *PCI Reading Program (PCI)* for students with severe disabilities as implemented in Brevard Public Schools (BPS) and Miami-Dade County Public Schools (M-DCPS). We report here on the results of the research that began in April 2007. Two earlier reports detail the results from the first two years (Toby, Ma, Jaciw, & Cabalo, 2008; Zacamy, Zheng, Jaciw, Ma, & Zhao, 2010).The primary question addressed by the study is whether students whose teachers have received the *PCI Reading Program* improve their reading skills as indicated by higher Sight word and Phonological assessment scores than students whose teachers have not had exposure to the program. The study began in Phase 1 (2007-2008 school year) as a randomized control trial (RCT) in BPS and M-DCPS to determine the efficacy of the *PCI Reading Program - Level One*. In the ensuing two years we analyzed longer-term trends by conducting quasi-experiments with teachers and students from the Phase 1 RCT serving as the *PCI* group and enlisting new teachers and students to serve as the control group.

This report presents the student outcome and classroom implementation findings and summarizes overall trends from all three phases of the study. In Phase 3, the small remaining sample size of students prevented researchers from being able to determine an impact on the Sight word or Phonological scores as was done in Phases 1 and 2 of the study. Instead, we investigated associations between student progress and years of participation in the program, while continuing to track the implementation of the program.

At the outset, the research questions were:

1. What is the impact of *PCI* on student achievement in sight word recognition and phonological awareness?
2. Is the impact of *PCI* different for different kinds of students or teachers?
3. How are reading programs implemented in participating classrooms?

Results from phases 1 and 2 will be summarized in this report. For the complete report of Phase 1, 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). For the complete report of Phase 2, please see *The Efficacy of PCI Education's* PCI Reading Program - Phase 2: *A Report of a Comparison Group Study in Brevard Public Schools and Miami-Dade Public Schools* (Zacamy, Zheng, Jaciw, Ma, & Zhao, 2010). A complete discussion of findings from Phase 3 is provided in this report.

Methods

Experimental Designs

During Phase 1 we conducted a randomized control trial (RCT) to test the effects of *PCI*. A coin was tossed to determine which teachers were assigned to teach *PCI* to their students and which were assigned to the control condition. Randomizing cases to conditions in this way ensures that the groups are statistically equivalent, which allows us to assess the impact of the program while limiting the biasing effects that would happen if teachers selected into the program in other ways, such as by volunteering. Our sample was composed of students with severe disabilities, and their teachers, from Brevard Public Schools and Miami-Dade County Public Schools. The experiment included 35 teachers (20 *PCI* and 15 comparison group teachers) and 128 students. In this first year of the study we looked solely for improvement in sight word reading since *Level One,* the only level of the *PCI* program available at that time, does not teach phonological skills.

At the end of the first year, Phase 1, the control teachers were trained in, and provided with, the *PCI* program. We integrated a quasi-experiment into the study in Phase 2 by recruiting additional teachers to serve as a comparison group. The process of up-taking experimental controls from Phase 1 into treatment in Phase 2 also allowed us to produce an extra-experimental estimate of the impact of *PCI* at the end of Phase 2 using differences in performance between the originally randomized groups at the end of each phase.

Phase 2 started with 22 teachers who agreed to use the *PCI Reading Program* with their students (*PCI* group). Eighteen of these teachers were from the original experiment (nine *PCI* and nine controls). We also recruited 17 teachers who had never used the *PCI* program to serve as the comparison group[1].

Phase 3 began with 20 teachers agreeing to use the *PCI Reading Program* with their students (*PCI* group) and 7 teachers who had never used the *PCI* program consenting to serve as a comparison group. Of the 27 teachers, 21 had participated in Phase 2 and 15 of those teachers had participated from the study's inception in Phase 1. In Phase 3 we were also prepared to conduct a quasi-experiment but the sample size at that point proved too small to support a statistical comparison.

Our power analysis calculations were based on an expectation of a fairly substantial impact. In Phase 1 of the experiment we decided to set the "minimum detectable effect size" (MDES) at 14 percentile points or, in terms of the standard deviation units, an effect size of 0.35[2]. For Phase 2 we selected the same value for the MDES. We estimated that 46 teachers (23 in each condition) would constitute a sufficiently large sample to detect an impact of at least 0.25 standard deviation units. However, for Phase 2 researchers were only able to recruit 34 teachers (21 *PCI* teachers and 13 comparison teachers.) Given this reduced sample size, we recalculated our MDES to 0.51. The experiment was 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 had an adequate sample size.

With our sample of 27 teachers in Phase 3 of this study, we expected to be able to detect an effect size of at least 0.56. However, by the end of Phase 3 we did not have a sufficient sample of students to conduct an impact analysis[3].

---

[1] The general approach that we used to maintain and recruit teachers and students in the study was to track students from an earlier phase (who were either part of the *PCI* group from the beginning or who served as controls or comparison cases) into their new classes and invite the teachers of those students to participate in the program. At the same time we would bring in a new set of cases to serve as the comparison group. For example, for Phase 2 of the study, researchers provided district staff members with the names of teachers and students involved during Phase 1. The districts then provided researchers with information about each student's school and teacher for 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. This same process was then repeated to select teachers for Phase 3 of the study.

[2] This is the smallest effect size that can be detected given specific tolerances for drawing false-positive and false-negative conclusions. Using this value and assumptions about other parameter values in the study we can determine the sample sizes necessary to detect the signal (the impact) amidst the noise (random variation).

[3] In Phase 1 teachers were randomized to conditions and, assuming students did not select into classes with knowledge of the assignment status of the class (e.g., make special efforts to be enrolled in *PCI* classes,) we start with statistically equivalent groups in that year. The two-year 'extra-experimental' impact estimate affords a level of protection against bias provided there is not a large degree of attrition of students who were originally randomized and provided that teachers are not selective in choosing to work with students depending on whether

Approach to the Quasi-experiment

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 that 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. Ideally the covariates used for matching would be collected at the same time in both conditions. Because we could not do so, there is some possibility of time exacerbating the differences between the groups being compared and limiting the quality of the matches. In spite of this possibility we were able to establish reasonable matches within strata as described below.

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 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 that had propensity scores outside the range of the propensity scores for the program cases. This led to the removal of several comparison cases for which matches had never been made. We divided the propensity scores for the program group into quartiles and then, within each quartile, checked for balance between the program and comparison cases simultaneously on the covariates used to calculate the propensity scores.[4] Balance was achieved within each quartile (although 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 that 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 that had propensity scores outside the range of the program 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

---

they received *PCI* in Phase 1 or not. By the end of Phase 3 very few students from the first year remained in the study, which limited the possibility of calculating a three year 'extra-experimental' impact estimate with precision and without statistical bias.

[4] 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.

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 noted by Shadish et al. (2006).

PCI Reading Program Levels One and Two

The program we are evaluating in 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

The trainings were similar in all three phases, except that during Phase 1 only *Level One* of the program was available and all teachers were new to the program. Here we focus on the training that took place at the beginning of Phase 3.

In Phase 3, *PCI* teachers in both districts were invited to participate in sessions to familiarize them with the *PCI Reading Program* as well as the research study. Comparison teachers were not invited to attend any part of the training in either district. Trainings for the two districts occurred at separate times and locations. In both districts, the first half (morning session) of the training was focused on introducing and preparing new teachers to use *PCI.* In M-DCPS one new teacher attended this session. In BPS, four teachers new to *PCI* attended. Also in attendance at BPS was a teacher who was not part of this study. That teacher's students had not previously received *PCI* instruction and would therefore have been designated to participate in the comparison group. However, this teacher instead selected to use *PCI* on her own and was therefore ineligible for participation in the study. In the afternoon, the newly-trained *PCI* teachers were joined by teachers who had implemented *PCI* during Phase 1 or Phase 2. In M-DCPS five teachers who had previously taught *PCI* joined for the afternoon session. Four teachers who had been participants in Phase 1 or 2 and continued to participate in Phase 3 were absent for the afternoon training. In BPS, four Phase 2 *PCI* teachers joined in the afternoon, and two Phase 2 *PCI* teachers were unable to attend. Also in attendance were two district support staff members and the district point of contact.

The training for the *PCI Reading Program*, 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 point of contact (POC) from PCI Education. The morning sessions focused largely on *Level One* of the program but also tied in *Level Two* and its relationship to *Level One*. During the training Haney introduced teachers to the previous research, as well as the literature and rationale behind *PCI*. In both districts, the training began with a theoretical and pedagogical overview of *PCI*. Haney explained the student prerequisites for starting each level and for moving forward within each level. In addition, she emphasized that the program should be implemented with a high level of fidelity to the Teacher's Guide. Throughout the training, Haney modeled various parts of the lesson cycle and then allowed teachers to practice with each other. When the teachers arrived in the afternoon, the session took a more informal direction, lending itself to question/answer and discussion formats. Teachers were invited to share both challenges and successes. In this session, teachers were also made aware of program updates and asked if they needed any additional materials. The training ended with a review of participant responsibilities and the research framework.

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. Nonreaders 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 was not yet released during the time of this research 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, repetition of the previous steps, and then review, assessment, and 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 students.

Expectations for Implementation

Expectations for implementation were discussed during the individual district trainings. *PCI* teachers are expected to use *PCI* as their core reading program for all appropriate students but the *PCI* trainer agreed that the use of supplemental materials was allowed. Teachers are expected to follow the curriculum and lesson cycles directly, as outlined in the Teacher's Guide. At a minimum, each student should receive 20 minutes of *PCI* instruction per day in order to comply with the publisher's definition of minimum acceptable implementation. According to *PCI*, ideal implementation is considered to be about 45 minutes per day for each student.

District Materials

In this study we compare *PCI* instruction to "business as usual." To gain an understanding of the existing reading materials in each district, researchers included questions on the teacher background form regarding the materials teachers had used during their prior school year. Three of the *PCI* teachers who had implemented *PCI* during Phases 1 and/or 2 indicated that they used supplemental reading programs in addition to using *PC*I, including Houghton-Mifflin, The Letter People, and Steck-Vaughan Reading. Teachers in the comparison group reported utilizing a variety of materials including Houghton-Mifflin, Reading Milestones, Edmark, Sonday, News-2-You, and other teacher developed materials.

Schedule of Major Milestones

Table 1 lists the major project milestones and associated dates.

Table 1. Research Milestones: Phase 1

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

Table 2 lists the major project milestones and associated dates. Planning for Phase 2 began in May 2008.

Table 2. 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

Table 3 below provides a broad timeline of research milestones from the beginning of Phase 3 to its completion.

Table 3. Research Milestones: Phase 3

Date	Milestone
August – December 2009	Recruitment of teachers
October – November 2009	Question-and-answer sessions and training
November 2009	Administration of pre-assessments and initiation of monthly web surveys
March 2010	Classroom observations
May 2010	Administration of post-assessments and completion of data collection

Participant Recruitment

District Identification

At the beginning of Phase 2, Empirical Education researchers contacted the person who had been designated as *PCI*'s Point of Contact (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 2-5. At the beginning of each school year of the study, the POCs identified eligible teachers who met the criteria to participate.

Teacher Identification

The way participants for a study are chosen largely determines how widely the results can be generalized. For example, where teachers are chosen because of exceptional characteristics, it may be difficult to find comparable teachers.

Teachers of students with supported level disabilities in grades 3-8 from the Miami-Dade and Brevard districts were identified by district staff and invited to participate in the study. Although in Phase 1 of the study the teachers were randomly assigned to the *PCI* group or the comparison group, this was not the case in Phases 2 and 3. All teachers who had participated Phase 1 were invited to participate in Phase 2 of the study as part of the *PCI* group—the group using the *PCI* program. All teachers from Phase 2 were invited to continue participation and retain their group assignment (*PCI* or comparison) for Phase 3. During Phases 2 and 3, most teachers new to the study were placed in the comparison group. However, exceptions occurred in cases where a student who had previously participated in the study as part of the *PCI* group was assigned to a teacher who had not previously participated in the research study. In these cases, the teachers were recruited to participate in the *PCI* group. Comparison teachers in Phase 2 and 3 agreed to continue to use their usual district materials until the 2010-2011 school year, at which time they would 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, to some extent we can control for the effects of these differences statistically. The Phase 1 results, being based on a randomized experiment, are especially authoritative. The validity of the Phase 2 outcomes is strengthened by the fact that we use two different methods that are grounded on different assumptions to corroborate the results.

Teacher Recruitment

For Phase 2 of the study, researchers provided district staff members with the names of teachers and students involved during Phase 1. The districts then provided researchers with information about each student's school and teacher for 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. This same process was then repeated to select teachers for Phase 3 of the study.

Because the next phase of recruiting happened internally, with the district POCs contacting all new 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.

To participate, teachers had to meet a set of criteria. The first criterion was that teachers must teach students with supported level disabilities, since the program is designed for students with intellectual disabilities and autism. The second criterion was that teachers must teach students who are in 3rd-8th grade. Finally, teachers must teach a self-contained reading block. This final requirement for participation was designed to ensure a measurable framework for implementation time that would be comparable across grades and various classroom settings.

All teachers who had been identified as potential study participants were sent an informational flier briefly describing the study and a participant information packet which included the following elements:

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

Researchers also hosted voluntary telephone question-and-answer sessions for all interested teachers. These sessions provided a format for the researchers to describe the specifics of participation as well as to answer potential participants' questions and address their concerns. Each district was offered two different dates/times to call in. Three teachers from BPS called in to the first session and no BPS teachers called in to the second session. No teachers from M-DCPS called in to either session.

Student Identification

Within the study classrooms, not all students are appropriate candidates for the *PCI* program. In addition to the recommendation that *PCI Level One* 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 when determining which students to include in the research study. In order for students to qualify for *Level Two* of the program, they must first have mastered the 140 sight words in *Level One*.

Site Descriptions

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 program. The results should not be considered to apply to school districts with practices and populations different from those in this experiment.

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 536, 357 in 2009 (U.S. Census Bureau, 2011).

BPS has 123 schools serving pre-kindergarten through grade 12. The total enrollment is 71,866 students (Florida Department of Education, 2011). Table 4 provides information about the entire district.

Table 4. Demographics of Brevard Public Schools

Brevard Public Schools	
Total schools[a]	123
Total teachers[b]	4,820
Grade structure	PK-12
Student enrollment[c]	71,866
Percent of students designated as:	
Disabled	17.4%
Orthopedically Impaired	0.005%
Speech Impaired	19.7%
Language Impaired	12.7%
Deaf or Hard of Hearing	0.005%
Visually Impaired	0.1%
Emotional/Behavioral Disabilities	5.5%
Specific Learning Disabled	42.0%
Hospital/Homebound	0.2%
Dual Sensory Impaired	<.001%
Autism Spectrum Disorder	4.0%
Traumatic Brain Injured	<0.001%
Developmentally Delayed	5.1%
Established Conditions	0.1%
Other Health Impaired	4.7%
Intellectual Disabilities	5.3%
English language learner[d]	2.4%
White[c]	64.5%
Black[c]	14.1%
Hispanic[c]	12.4%
Asian/Pacific Islander[c]	2.0%
American Indian/Native Alaskan[c]	0.2%
Multi racial[c]	6.8%

[a] Florida Department of Education, 11/2010
[b] Florida Department of Education, 2009
[c] Florida Department of Education, 2010
[d] Florida Department of Education, 2009-2010

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 2,500,625 in 2009 (U.S. Census Bureau, 2011).

M-DCPS has 515 schools serving pre-kindergarten through grade 12. The district's total enrollment is 347,406 students (Florida Department of Education, 2011). Table 5 provides information about the entire district.

Table 5. Demographics of Miami-Dade County Public Schools

Miami-Dade County Public Schools	
Total schools	515
Total teachers	21,139
Grade structure	PK-12
Student enrollment	347,406
Percent of students designated as:[a]	
Disabled	10.9%
Orthopedically Impaired	1.1%
Speech Impaired	7.5%
Language Impaired	1.9%
Deaf or Hard of Hearing	1.1%
Visually Impaired	0.4%
Emotional/Behavioral Disabilities	9.9%
Specific Learning Disabled	51.8%
Hospital/Homebound	0.9%
Dual Sensory Impaired	<0.001%
Autism Spectrum Disorder	6.4%
Traumatic Brain Injured	0.2%
Developmentally Delayed	4.2%
Established Conditions	0.2%
Other Health Impaired	7.1%
Intellectual Disabilities	7.5%
English language learner[b]	17.1%
White[c]	8.6%
Black[c]	24.5%
Hispanic[c]	65.1%
Asian/Pacific Islander[c]	1.2%
American Indian/Native Alaskan[c]	0.1%
Multi racial[c]	0.5%

[a] Florida Department of Education, 2009
[b] Florida Department of Education, 2009-2010
[c] Florida Department of Education, 2010

Data Sources and Collection

This research employs a mixed methods approach through which we measure and document *PCI* implementation to provide qualitative and quantitative descriptions of the program. 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.

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 ID
- Gender
- National School Lunch Program status (proxy for socio-economic status)
- Ethnicity
- English learner status
- Date of birth
- Grade
- Classroom teacher
- School the student attends
- Disability/Eligibility codes

Due to the parental consent requirements in M-DCPS, researchers only received data for students who 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).

In Phase 2 Empirical Education received these data from M-DCPS and BPS in January 2009. In Phase 3, data were obtained from M-DCPS and BPS in January and February 2010.

Achievement Measures

The primary outcome measures are student assessment scores on the Sight word assessment and Phonological assessment. In this section, we outline the development of the Sight word and Phonological pre- and post-assessments

Sight Word Pre- and Post-Intervention Assessment

Prior to Phase 1, a development specialist took the following steps to determine the appropriate words for both the Sight word pre- and posttests.

1. Selected only words that are taught in both the *PCI* and Edmark reading programs. Edmark is a reading program that was initially thought to be used in a majority of participating classrooms.
2. Used the Evaluation Description Language (EDL) Reading Core Vocabulary Cumulative list to determine the reading levels of each word. This was important so that each of the two 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.

Prior to Phase 2 the Sight word assessment list used in Phase 1, along with the list of *Level One* and *Level Two* words from the program, were sent to an Associate Professor from Florida State University's College of Education and the Center for Reading Research. This professor then 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* that would likely appear in comparison reading programs. A statistician from Empirical Education computed correlation coefficients for the Sight word and Phonological assessments, as well as correlations between the different version of the Sight word assessments (pre- and post-).

For Phase 3, a psychometrician from Empirical Education developed Sight word pre- and post-assessments that would allow us to link Phase 2 and Phase 3 assessments, as well as Phase 1 and Phase 3 assessments (and therefore Phase 1 and Phase 2). The psychometrician used the following guidelines, in order of priority, when selecting words for both the pre- and post-assessments.

1. Each test in Phase 1 through 3 (with the exception of the Phase 1 pre-assessment developed by the consultant) includes five core items of comparatively easy words from *Level One*. Three of those items were retained for future test administrations after Phase 3. However, future phases of the study have been canceled.
2. Each test includes at least six secondary items that were present at least three times during Phases 1-3.
3. Beginning with Phase 3, tests include a few *Level Two* words to link to future assessments. For the Phase 3 pre-assessment, only two *Level Two* words are included because the majority of participating students had not reached *Level Two* by the end of Phase 2. The number of *Level Two* words included in the post-assessment is determined based on the reading levels of the participating students in Phase 3.
4. Each secondary item has an equal chance of being included in either the pre- or post-assessment.
5. Each assessment is balanced with regard to word difficulty, as determined by the Brown Corpus list. Research shows that word frequency and word difficulty are closely related (Breland, 1996). The frequency of use is a valid measure of word difficulty that can be utilized in vocabulary test construction and translation.
6. Overall, the assessments cover a broad range of words from the *PCI* word list.

Each sight word assessment consists of 20 items and items given in the assessments are randomly ordered. An exception is made for the first item on the test in order to allow students to begin the assessment with an easy item.

Phonological Pre- and Post-Intervention Assessment

In addition to the Sight word assessment, a Phonological assessment was developed by an independent consultant during Phase 1. It was not used as an outcome measure in Phase 1 because teachers and students were only exposed to *Level One* of the program, which does not include phonics instruction. *Level Two*, however, introduces phonics instruction, and therefore, during Phases 2-3, it became important to track improvement of phonics skills.

The same 15 Phonological assessment items used as a pre-assessment in Phase 1 were used as a pre-assessment in Phase 2. However, upon deciding that a phonological assessment would be used as an additional outcome measure in Phase 2, researchers solicited recommendations from the Associate Professor from Florida State University who had provided words for the Sight word pre- and post-assessments. After reviewing the cost, testing time, and appropriateness of the recommended assessments, researchers selected the DIBELS Initial Sound Fluency as the Phonological post-assessment measure (16 items). 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. This request was approved for our research with this population of students.

Pre- and Post-Assessment Procedures

Researchers provided teachers with a pre-assessment packet at the beginning of the school year and mailed post-assessment packets in early May. For both districts, the pre-assessments were administered between November 2009 and January 2010, and all post-assessments were conducted between May and June 2010.

While teachers who had been involved in previous years of the research study had experience administering the Sight word and Phonological pre-assessments, the other teachers most likely had some experience in administering sight word assessments but were not already familiar with the specific administration of the research study's assessments. Therefore, researchers provided detailed instructions in the assessment packets and encouraged teachers to contact the researchers if they had any questions. Specifically, the packets contained a list of student prerequisites to determine which students should be assessed and included in the study. In addition, the handouts reiterated to both *PCI* and comparison teachers that the pre-assessment was to be administered before instruction in the program began. Teachers also received the following materials in the pre- assessment packets.

- a cover letter describing the packet contents and directions for returning the assessment items (one per packet)
- Sight word assessment administration, scoring guide and word card set (one per packet)
- Sight word assessment scoring sheet (one per student)
- Phonological administration, scoring guide, and picture card set (one per packet)
- Phonological assessment/DIBELS Initial Sound Fluency: short form directions/progress monitoring sheet (1 per student)
- pre-stamped envelope for returning materials (1 per packet)

Teachers received the following materials in the 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; *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)
- Sight word assessment scoring sheet (one per student)
- Phonological administration, scoring guide, and picture card set (one per packet)
- Phonological assessment/DIBELS Initial Sound Fluency: short form directions/progress monitoring sheet (1 per student)
- pre-stamped envelope for returning materials (1 per packet)

Assessment Administration

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. Researchers followed up with survey questions which asked how teachers had determined which students to assess.

Teachers were instructed to first fax and then to mail completed pre-assessments to Empirical Education in postage-paid envelopes. The same process was used to distribute and obtain the post-assessment measures. For Phase 3 we received completed assessments from all but one participating teacher.

Methods Used to Investigate the Program Implementation

In addition to assessment scores, researchers collected data on program implementation over the entire study, beginning with the teacher recruitment phase and ending with the academic calendar of the district in each of the three study years. 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. The majority of our implementation data come from survey questions. Data from other sources are mostly used to triangulate with the survey data.

Survey Schedule

Surveys were deployed to both *PCI* and comparison group teachers beginning in November of each year and continued on a monthly basis through May. The survey schedule and response rates were similar during all three phases. Table 6 provides this information for Phase 3.

Table 6. Survey Schedule

Survey	Deployment	Response rate
Survey 1	November 20, 2009	100%
Survey 2	December 11, 2009	96%
Survey 3	January 15, 2010	96%
Survey 4	February 19, 2010	96%
Survey 5	March 19, 2010	96%
Survey 6	April 9, 2010	96%
Survey 7	April 30, 2010	96%
Survey 8	May 14, 2010	93%

Note. (n = 27)

Classroom Observations

In general, observational data are used to inform the description of the learning environment, instructional strategies employed by the teachers, and student engagement. These data are minimally coded.

1) Classroom observations took place in both districts during the spring for each of the study years. Our goal was to visit one-third of participating teachers and classrooms. Class selection for observations was based on "stratified convenience," that is, 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, as follows. In each district, we first selected schools containing multiple teacher participants in order to maximize the number of teachers observed within the allotted time.
2) 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 were in their first, second, and third year of *PCI* implementation.
3) 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.

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.

In Phase 3, at M-DCPS, we visited seven out of 12 schools and seven out of 14 classes. At BPS, we visited five out of 11 schools and six out of 13 classes. We had scheduled one additional observation with a Miami teacher; however, since that teacher's sole study student was absent on the day of the scheduled observation, no observation took place. In total, across both districts, we visited 12 out of 23 schools and 13 out of the 27 classes in Phase 3. Of the 13 classes, six were elementary school classrooms, nine were *PCI* classrooms, and three teachers (one *PCI* and two controls) were new to the research study in Phase 3.

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 the types of curricula that 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 the observations in both districts as well.

Implementation Data, Collection, and Analyses

Table 7 lists the implementation areas of investigation for some or all of the phases of this study, the types of analyses conducted for each area, and the data sources. These components and the rationale for the analyses are explained below.

Table 7. 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 and 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 related to the different levels of *PCI* implementation	Teacher surveys
Impact of Instruction	Investigate whether *PCI* instruction is associated with the number of minutes of reading instruction students received as well as student engagement.	Teacher surveys

Teacher Background

This study collects teacher background data to provide a context for reading program implementation. Because recent literature correlates teaching experience and content knowledge with teacher quality (Amrein-Beardsley, 2006; The Center for Public Education, 2005), we conduct balance checks on teacher background data to establish comparability between the *PCI* and comparison groups. In addition, we planned to conduct a moderator analysis on teachers' years of experience as a special education teacher, if there was sufficient

variation between teachers' years of experience to do such an analysis. The data collected included:

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

Conditions for Implementation

In addition to contextual information regarding teacher background, it is critical to have information about the conditions under which the implementation takes place in order to understand the rest of the implementation and outcome data. We present the data using descriptive statistics, and we draw comparisons between the *PCI* and comparison groups.

Program Training

During the planning stages of this study, PCI Education expressed a particular interest 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

We also investigated whether participating teacher had 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. The surveys asked teachers in both groups to confirm that they had the necessary materials to fully implement their reading programs.

Description of Implementation

While the primary focus of this study across the three phases is on student outcomes, the results need to be understood within the context of classroom implementation. Here we present our rationale for exploring each indicator of program implementation.

Reading Materials Used

We begin our implementation results section by comparing the materials used by teachers in the *PCI* group to those used by the comparison teachers. On the background information form included in the consent packets, teachers in both conditions indicated that in the past they had 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 classrooms.

Teacher Satisfaction

We also queried teachers about their experiences with their reading materials. 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 has been 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 classrooms. *PCI* teachers were asked questions about student progress, classroom and instructional organization, student assessment, and program bonus materials.

- **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 materials that are supplementary for *Level One,* but that are 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

PCI education was also interested in learning whether student achievement would increase as instructional time with the *PCI Reading Program* increased. Therefore, with each survey we asked *PCI* teachers to report the amount of time students received *PCI* instruction during a given week. From teachers' answers to these questions, researchers planned to run a correlation between the amount of instructional time and student achievement outcomes.

Impact on Instruction

Reading Instruction Time

Researchers investigated whether use of the *PCI* program was associated with the number of minutes of reading instruction students received. In other words, researchers investigated whether teachers in *PCI* and comparison groups differed in the amount of reading instruction provided to participating students. 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 found a significant decrease in the number of minutes of *PCI* instruction reported during the Florida Alternative Assessment (FAA) testing period. Therefore, teachers were asked if instructional practices changed or reading instruction changed in any way in order to prepare for standardized testing. This decrease in minutes was observed again in Phase 2 and therefore re-incorporated into the Phase 3 survey questions.

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.

Teacher Responses on Implementation Data

Reports on percentages of teachers responding to specific survey questions were calculated by dividing the number of teachers who selected a specific response by the number of teachers that were asked, and provided a valid response to, that question. 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.

Formation of the Experimental Groups: Phase 1 and Phase 2

This section describes the samples that were used to determine the impacts for Phase 1 and 2.

Characteristics of the *PCI* and Comparison Groups in Phase 1

For Phase 1, the sample consists of teachers and students, where teachers have been randomly assigned to *PCI* or the control condition. We describe this sample as being formed initially through the random assignment but modified somewhat through attrition or loss of units at different points during the experiment for a variety of reasons. Ideally, by randomizing assignment into the two conditions, the groups should look the same in terms of important characteristics such as demographic composition, prior achievement, and teacher characteristics. In addition because we paired teachers, we can expect somewhat better balance than we would have if we hadn't first balanced them on these characteristics. However, by chance (as well as the imprecision of the pairing) the groups are never exactly balanced and may differ on important characteristics likely to affect the outcome. Furthermore, the loss of teachers and students during the experiment may result in imbalance in the sample. This would happen, for example, if teachers are more likely to drop out of the program than the control group due to additional burden.

Therefore in this section we inspect the distribution of teachers, classes, grades, and students, looking in particular at the balance between the *PCI* and control groups. For detailed information

on attrition, please reference *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)

Number of Units in the Sample and Attrition in Phase 1

Table 8 shows changes in the sample from the point at which the teachers were randomized to the point at which the posttests were received.

Table 8. Numbers of Units in the Experimental Groups and Attrition in Phase 1

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

Characteristics of the Initial Sample in Phase 1

In Table 9 we compare the composition of the control and *PCI* groups at the point we received the rosters. Some attrition did occur between the time of randomization and the point at which we received the rosters. Therefore it is useful to examine the groups that we received information on to determine whether initial random assignment and subsequent attrition up to the point of receiving roster information[5] resulted in a difference between conditions on background characteristics. The high *p* value associated with each of the tests indicates that the differences could easily be the result of chance.

[5] We report on attrition to this point because this was the stage at which we produced an interim report to PCI that included a description of the overall sample and losses of cases.

Table 9. Student Characteristics of Phase 1 Sample

	Control group	*PCI* group	Less than 5% chance of seeing this much imbalance
English proficient	75 (98.68%)	73 (94.80%)	No (p =.27)
Low socio-economic status	56 (73.68%)	60 (77.92%)	No (p =.82)
Mean Sight word pre-test score	4.26	5.73	No (p =.32)

Distribution by Grade in Phase 1

Table 10 shows the distribution by grade of the 150 eligible students for whom rosters and pretests were received.

Table 10. Distribution of Students by Grade in Phase 1

	Grade Level						
	3	**4**	**5**	**6**	**7**	**8**	**Total**
Control	10	8	1	17	19	18	**74**
PCI	10	12	12	11	16	16	**76**
Total	**20**	**20**	**13**	**28**	**35**	**34**	**150**

Characteristics of the *PCI* and Comparison Groups in Phase 2

For Phase 2, the *PCI* group is composed of teachers (and all of their students) who had participated in the study during Phase 1 or who received students who had *PCI* instruction during Phase 1. The comparison group consists of all other teachers who taught target students and who agreed to participate in the study. The data from the comparison group was used in the quasi-experimental analysis. The intent in selecting a comparison group is that it will be similar to the program group 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 an additional burden imposed on that particular group. Therefore, in this section we inspect the data for 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.) 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.

Number of Units in the Sample and Attrition in Phase 2

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 11 shows counts for the quasi-experimental analysis of the two-year impact. Table 12 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 11. Numbers of Units in the Quasi-Experimental Analysis of the Two-year Impact in Phase 2

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 is Sight word; PH is Phonological)	**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] *PCI* 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 *PCI* cases.

Table 12. Numbers of Units in the Extra-experimental Analysis of the Two-year Impact in Phase 2

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 is Sight word; PH is Phonological)	**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 in Phase 2

Table 13 shows some of the background characteristics of all qualified teachers and students (given in the last row of Table 11) used in the quasi-experimental analysis of the two-year impact for the Sight word assessment.[6]

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 overlap between conditions

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

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

Table 13. Characteristics of the Teachers and Students in Phase 2

	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

Note. Possible scores for Sight word pretest ranged between 0 and 20; possible scores for Phonological pretest ranged between 0 and 16.

Identification of Student Groups and Analysis Plan

Student Groups

Due to the study design and the criteria set for teacher participation, students began Phase 3 with different levels of exposure to the program, including some students who had been part of the initial randomized group in Phase 1. Students fall into eight groups based on their levels of exposure over the three year period (Phase 1 through Phase 3). Table 14 shows the patterns of exposure of the different student groups.

Table 14. Student Groups

Student group	Phase 1 (2007 - 2008)		Phase 2 (2008 - 2009)		Phase 3 (2009 - 2010)	
	PCI	Control	PCI	Comparison	PCI	Comparison
Group 1	X		X		X	
Group 2	-	-	X		X	
Group 3	-	-	-	-	X	
Group 4		X	X		X	
Group 5		X		X		X
Group 6	-	-		X	X	
Group 7	-	-		X		X
Group 8	-	-	-	-		X

Note. "X" indicates participation of the student group, and "-" indicates that the student group was not involved in that Phase.

Figure 1 further illustrates the student group patterns presented in Table 14. 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.

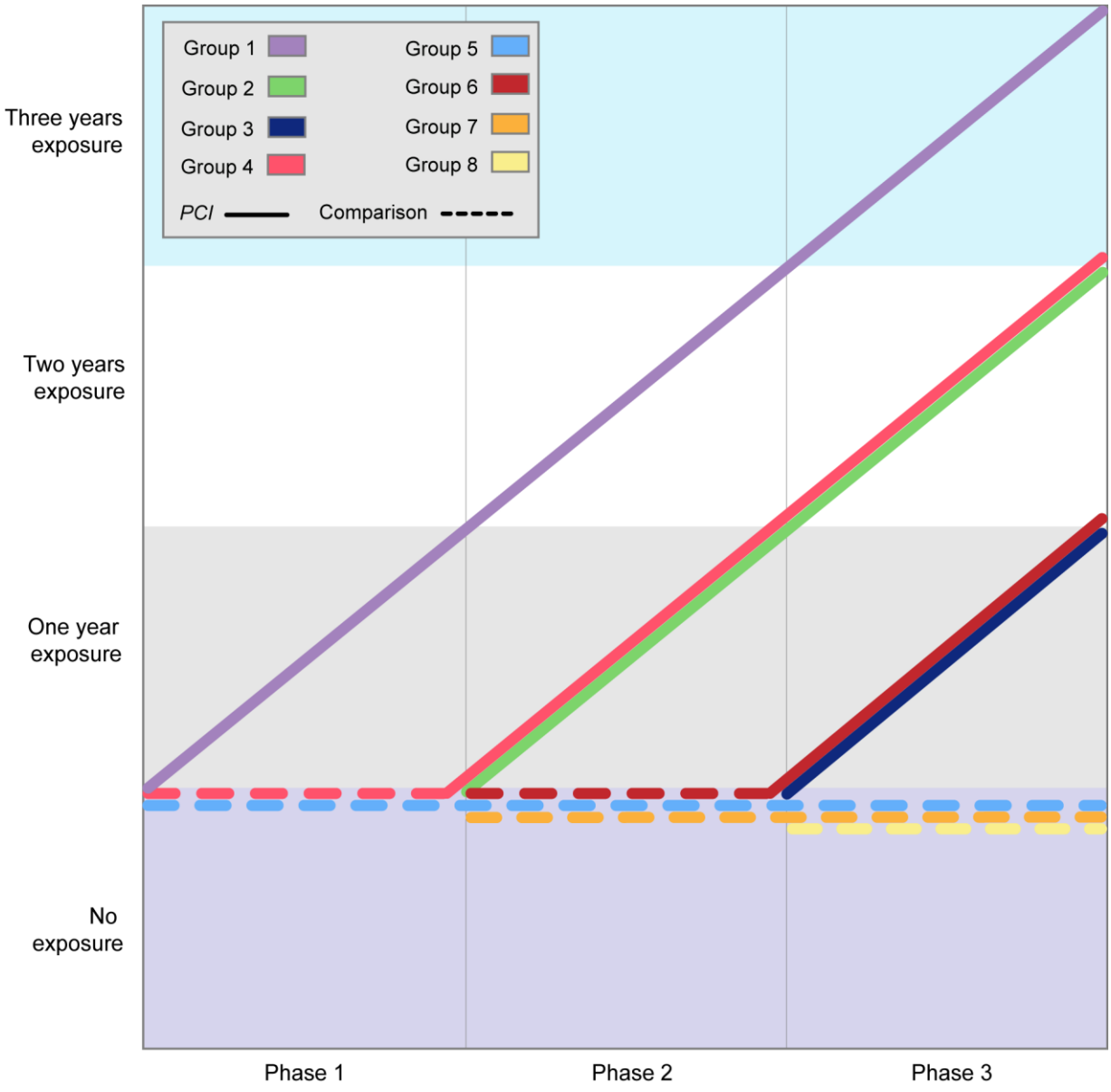

Figure 1. Illustration of Student Groups by Exposure to *PCI* during the Three Phases of the Study

Analysis Plan

Here we summarize the initial analysis plan for Phase 3 in order to better understand why we were unable to determine any impact of the *PCI* program in Phase 3. A primary purpose of the Phase 3 study was to estimate an impact of the intervention after three years. However, only six students remained in the treatment condition over the course of the three years of the study, which

precluded running this analysis. We realized after the first semester of Phase 3 that we would not be able to conduct a quasi-experimental 3-year impact analysis or an extra-experimental impact analysis that would be adequately powered and give conclusive results. Instead, we had hoped to estimate the 2-year impact of *PCI* using the quasi-experimental (QE) approach. In this way, different student groups would be compared to determine impacts of the program on sight word recognition as well as on phonemic awareness during the two year period (Phase 2 through Phase 3).

The QE 2-year impact analysis for the sight word outcome would have involved students who had used *PCI* for two years (i.e. Group 2 who were admitted to *PCI* at the start of Phase 2 and continued to use *PCI* in Phase 3, and Group 4 who were randomized to control in Phase 1 and exposed to *PCI* in Phases 2 and 3), as well as students who had never received exposure to *PCI* (i.e. Group 5 + Group 7 + Group 8). We did not plan to run a QE 1-year impact analysis on the sight word outcome, since the impact estimate was produced using the Randomized Control Trial (RCT) data from Phase 1, which would not be improved upon through replication with a QE analysis.

The QE 2-year impact analysis for the phonological outcome was to include the same students who had used *PCI* for two years. Moreover, we planned to conduct a QE 1-year impact analysis on the phonological outcome comparing students who have used *PCI* for one year in Phase 3 (i.e. Group 3 who were admitted to *PCI* at the start of Phase 3, and Group 6 who were admitted as comparison in Phase 2 and exposed to *PCI* in Phase 3), as well as students who had never received exposure to *PCI* (i.e. Group 5 + Group 7 + Group 8).

We would also have investigated whether *PCI* had a different effect on sight word recognition and phonemic awareness for specific subgroups:

- Students who scored lower on the Sight word and Phonological pre-assessments
- Students in lower or higher grades
- Students who have teachers with more experience teaching special education
- Autistic students
- Students who are English language learners

In addition to examining impacts and interactions where we anticipated effects, we may have conducted a limited number of exploratory analyses to better understand unexpected results. These may have included correlating implementation factors with outcomes, such as the number of minutes *PCI* teachers spend teaching reading.

Formation of the Experimental Groups: Phase 3

Characteristics of the *PCI* and Comparison Groups in Phase 3

This section describes the sample that we intended to use to determine the relationship between the *PCI Reading Program* and the measured outcomes.

The program group is composed of teachers and students who received *PCI* in Phase 1 or Phase *2*. The comparison group consists of all other teachers of target students. The comparison group was to be used in the quasi-experimental analysis. Normally, we would inspect the teacher and student data and check whether the *PCI* and comparison groups are balanced on important characteristics. However, because we aren't running any analyses to determine any effect on the performances of these two groups, balance checks are no longer an important goal. Instead, we look at the attrition that prohibited us from conducting these analyses.

Number of Units in the Sample and Attrition in Phase 3

Table 15 shows 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 15 shows the number of consented students we have with 1-year Phonological scores and 1-year and 2-year Sight word assessment scores.

Table 15. Number of Units in the Sample and Attrition in the Analysis of the Two-year Impact in Phase 3

Event	Comparison		*PCI*	
	No. of teachers	No. of students	No. of teachers	No. of students
1 year Phonological assessment total	**7**	**26**	**17**	**57**
1 year Phonological assessment with posttest	7	22	17	48
2 year Phonological assessment total	**7**	**26**	**13**	**34**
2 year Phonological assessment with posttest	7	22	13	30
2 year Sight word assessment Total	**6**	**26**	**13**	**34**
2 year Sight word assessment with posttest	7	20	13	30

Characteristics of Teachers and Students in Phase 3

Table 16 shows background characteristics of all qualified teachers and students. We see that there is balance between conditions in teachers' average years of special education teaching experience and in certain student characteristics (number autistic, gender, and socio-economical status). Researchers did find a significant imbalance in ethnicity between assignment groups; however, this imbalance does not pose a concern since we were not able to conduct an impact analysis between the groups.

Table 16. Characteristics of the Teachers and Students in Phase 3

	Comparison group	*PCI* group	Is the imbalance significant?
Teachers			
Fewer than four years Special Education teaching experience	1 (14.29%)	4 (21.05%)	No
Students			
Black	2 (7.69%)	27 (27.27%)	Yes
Male	18 (69.23%)	59 (59.60%)	No
National School Lunch program	17 (65.38%)	75 (75.76%)	No
Autistic	5 (19.23%)	14 (14.14%)	No

Table 17 presents the number of students appearing across each of the three phases, as well as the breakdown of Autistic and non-Autistic students, for both the comparison and control groups. The number of students remaining in the study for three complete phases is very small compared to the number of students in one phase.

Table 17. Characteristics of the Teachers and Students

Groups	Total no. of students	No. of autistic students	No. of non-autistic students
Comparison students in one phase	13	3	10
Comparison students in two phases	11	2	9
Comparison students in three phases	2	0	2
***PCI* students in one phase**	57	6	51
***PCI* students in two phases**	34	6	28
***PCI* students in three phases**	8	2	6
Total	**125**	**19**	**106**

Descriptive Characteristics of Teachers and Classroom Descriptions in Phase 3

Here we report survey data on additional characteristics of all study teachers and classrooms. .

Teacher Background in Phase 3

At the beginning of the 2009-10 academic year, teachers responded to survey questions regarding their teaching experience and educational background. Provided in Table 18 and

Table 19 are the responses regarding years of teaching experience for each condition, as well as the average number of years of experience teaching Special Education for each condition. Teachers across both assignment groups have generally similar characteristics.

Table 18. Distribution of Years Teaching Experience in Phase 3

	No. of teachers		
Condition	**0 to 3 years**	**4 or more years**	**Totals**
Comparison	1	6	7
PCI	3	16	19
Totals	4	22	26

Statistical test	***p* value**
Fisher's exact test	1.00

Table 19. Distribution of Years Teaching Special Education in Phase 3

	No. of teachers		
Condition	**0 to 3 years**	**4 or more years**	**Totals**
Comparison	1	6	7
PCI	4	15	19
Totals	5	21	26

Statistical test	***p* value**
Fisher's exact test	1.00

Table 20 provides data on teacher credentialing. All teachers in both assignment groups held a regular/standard teaching certificate. Four of the eight teachers (50%) that chose "other" specified that they also had an English for Speakers of Other Languages (ESOL) endorsement.

Table 20. Teacher Credentialing and Certification in Phase 3

	Regular/ standard	NBPTS[a]	Florida's Autism Endorsement	Speech-Language Pathology	Specific certificates for teaching bilingual, multicultural, or limited English	Other
Comparison (*n* = 7)	7 (100.0%)	2 (28.6%)	1 (14.3%)	1 (14.3%)	2 (28.6%)	0 (0.0%)
***PCI* (*n* = 20)**	20 (100.0%)	2 (10.0%)	2 (10.0%)	1 (5.0%)	3 (15.0%)	8 (40.0%)

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

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

Table 21 presents teachers survey responses regarding their highest level of education completed. A larger percentage of teachers in the comparison group had obtained a Master's degree (5 out of 7 or 71%) than among teachers in the *PCI* group (9 out of 20 or 45%).

Table 21. Highest Level of Education Completed in Phase 3

	Bachelor's degree	Master's degree	Education specialist or professional diploma	Doctorate or first professional degree
Comparison (*n* =7)	2 (28.6%)	5 (71.4%)	0 (0.0%)	0 (0.0%)
***PCI* (*n* = 20)**	11 (55.0%)	9 (45.0%)	0 (0.0%)	0 (0.0%)

Classroom Description in Phase 3

Teachers were asked to select all the designations of disabilities that exist within their participating classroom. One teacher in the *PCI* group reported that their classroom included an orthopedically impaired student, which is represented in the "other" category in Table 22.

Table 22. Description of Student Designations in Phase 3

	Intellectual disabilities at the supported level	EMH	TMH	PMH	ASD	Varying exceptionalities	Other
Comparison (*n* = 7)	5 (71.4%)	1 (14.3%)	3 (42.9%)	0 (0.0%)	3 (42.9%)	0 (0.0%)	0 (0.0%)
***PCI* (*n* = 20)**	12 (60.0%)	2 (10.0%)	8 (40.0%)	0 (0.0%)	5 (25.0%)	0 (0.0%)	1 (5.0%)

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

Analysis and Reporting on the Impact of *PCI Reading Program*

This section describes the statistical analysis and reporting. 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 implementation of the program, students in *PCI* classrooms would have higher Sight word reading and Phonological assessment scores than students in comparison classrooms. To appropriately estimate this difference in the quasi-experimental analysis, our equation contains a term to distinguish the records for the *PCI* students from those for control students, as well as terms for other important factors, called covariates, associated with characteristics of teachers or students, which we expect to make a difference in the outcomes. 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 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 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 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. Assuming that the equivalence established through randomization in Phase 1 is maintained through Phase 2, the pretest serves to adjust for chance imbalances rather than systematic differences.

Moderators at the Student and Teacher Level

In addition to adding covariates to the model, we consider whether there is a difference in the effect of the program 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) in the outcome by including an interaction term in the statistical equation. This term multiplies together the variable that indicates whether the student is in the program 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.

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 a measure of the variability in the outcome.

The amount of variability is also called the "standard deviation" (technically, the square root of the variance) and can be thought of as the average distance of all the individual scores from the average score. 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. When possible, we also report the effect size of the difference after adjusting for the effects of the pretest score and other fixed effects, since that adjustment provides a more precise estimate of the program effect by compensating for average differences on these covariates between the program and comparison groups. 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.

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 program 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

The following sections provide a summary of the implementation results from Phase 1 and Phase 2 as well as detailed implementation results from Phase 3. The summaries of Phases 1 and 2 highlight trends and themes that we have observed across all phases of the study and recapitulate significant implementation findings from the first two phases. The implementation results for Phase 3 will also address these trends in greater detail as well as other measures of implementation not reported in the summaries of the first two phases. Across the three phases main themes include high teacher satisfaction and student engagement as well as slower student progress than expected, which we examine closely in Phases 2 and 3.

Phase 1 Implementation Results Summary

The conditions for implementation of the *PCI* program in Phase 1 appeared to be quite good. All teachers in the *PCI* program received training, while few teachers in the control group reported having ever received training in their reading program. Eighty-six percent of the *PCI* teachers had all the

materials they needed at the time of the training. The *PCI* teachers were offered support from PCI although no teachers appeared to take advantage of this option. However participants did report going to each other and district supervisors for help.

As seen in Figure 2, 65% of *PCI* teachers reported that they were very satisfied with their reading program, as compared to only 21% of control teacher. Not one *PCI* teacher reported dissatisfaction or even a neutral opinion of the program.

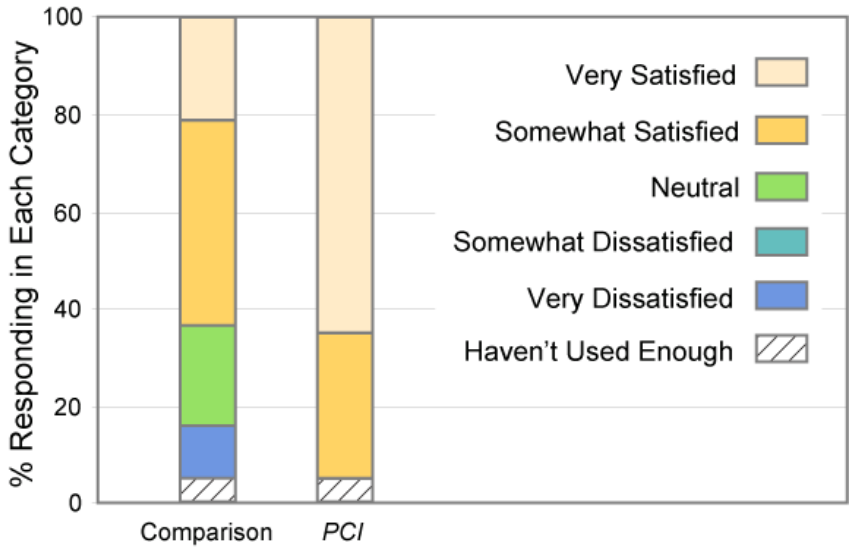

Figure 2. Teacher Satisfaction with Reading Program in Phase 1[7]

Teachers were also asked to rate student level of engagement while participating in various aspects of the program. Students would be considered fully engaged if they displayed consistent on-task behavior. Teachers observed high levels of engagement and enjoyment in their students overall. Certain activities, such as reading books and the Trace and Read Workbooks, ranked very high in terms of student engagement, while other activities such as the Word Game received more moderate rankings.

[7] *n* = 19 for Comparison; *n* = 20 for *PCI*

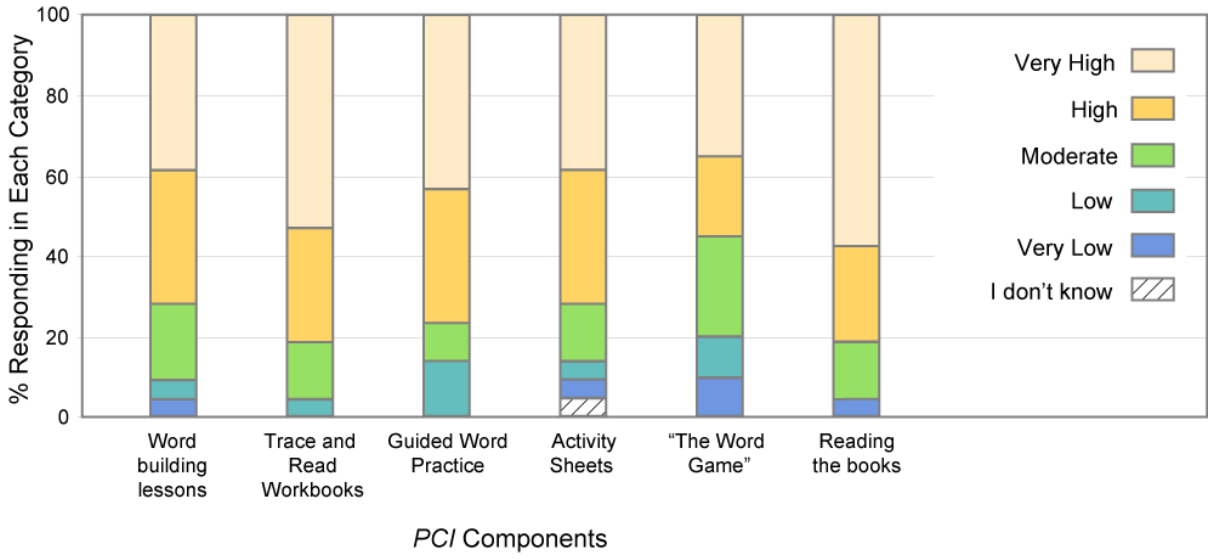

Figure 3. Teacher Report on Levels of Student Engagement in Phase 1 with *PCI* Components[8]

PCI teachers appeared to fully implement the program, although at minimum rather than optimal levels. The publisher defines ideal implementation as 45 minutes per student per day and the minimum acceptable implementation as 20 minutes per student per day. Through the course of the academic year, teachers in the *PCI* group recorded accomplishing the minimum amount of instruction as specified by the publisher for implementation compliance. Almost all of these teachers supplemented the *PCI* program with other curricular materials. While students did not progress as far in the program as initially expected, teachers did generally follow the lesson cycle as specified by the publisher. While some teachers reported difficulty in finding time for the amount of individualized instruction necessary, opinions of the program itself were high. Teachers in the *PCI* program reported higher levels of student enjoyment and engagement, as well as general satisfaction with the program, than did teachers in the control group. Teachers were excited to see their students reading books and retaining words they learned through the program.

Phase 2 Implementation Results Summary

The conditions for implementation for *PCI* in Phase 2 again appeared to be good. Twenty of 21 (95%) teachers in the *PCI* program received training from PCI Education or a district official, while only five of the 13 (39%) teachers in the control group reported ever receiving training for their reading programs. When surveyed in January, 100% of the *PCI* teachers reported having all the materials they needed.

At the end of the academic year, teachers in both assignment groups were asked about their satisfaction with their reading curricula. As shown in Figure 4, 12 out of 19 *Level One* teachers (63%) and four out of six *Level Two* teachers (67%) reported that they were very satisfied with the program,

[8] *n* = 21 for all components except for the Word Game. One teacher reported "N/A" for each of the components except for the Word Game. Two teachers reported "N/A" for the Word Game.

as compared to only one comparison teacher. As in Phase 1, not a single *PCI* teacher reported dissatisfaction with the program.

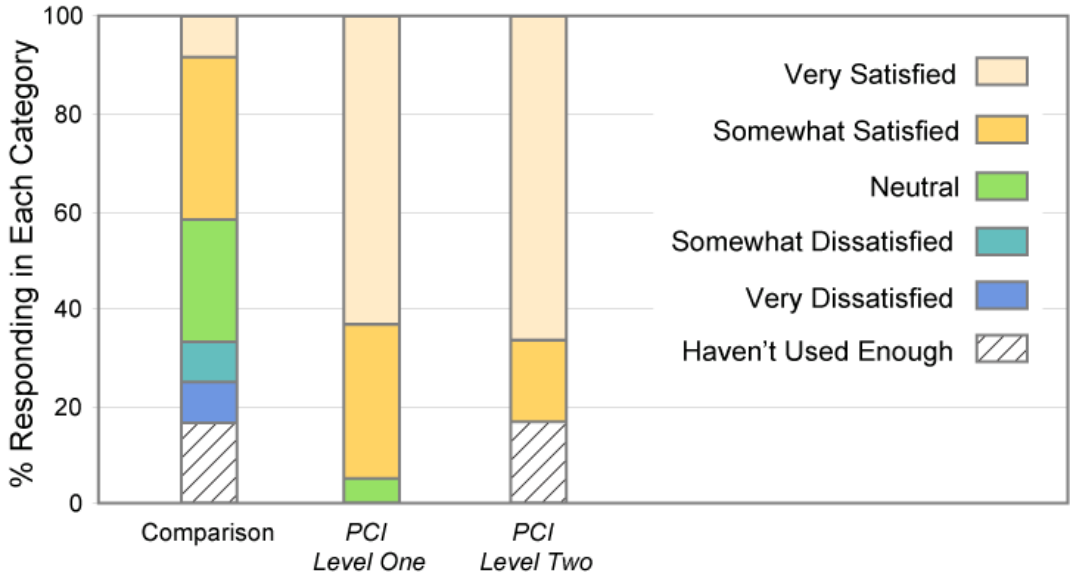

Figure 4. Teacher Satisfaction with Reading Program in Phase 2

Researchers asked teachers in both groups to rate the average level of student engagement with their reading programs. As in Phase 1, teachers were instructed to consider students 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. *PCI* teachers were also asked to rate student level of engagement while participating in various aspects of the program. As shown in Figure 5 and Figure 6, 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.

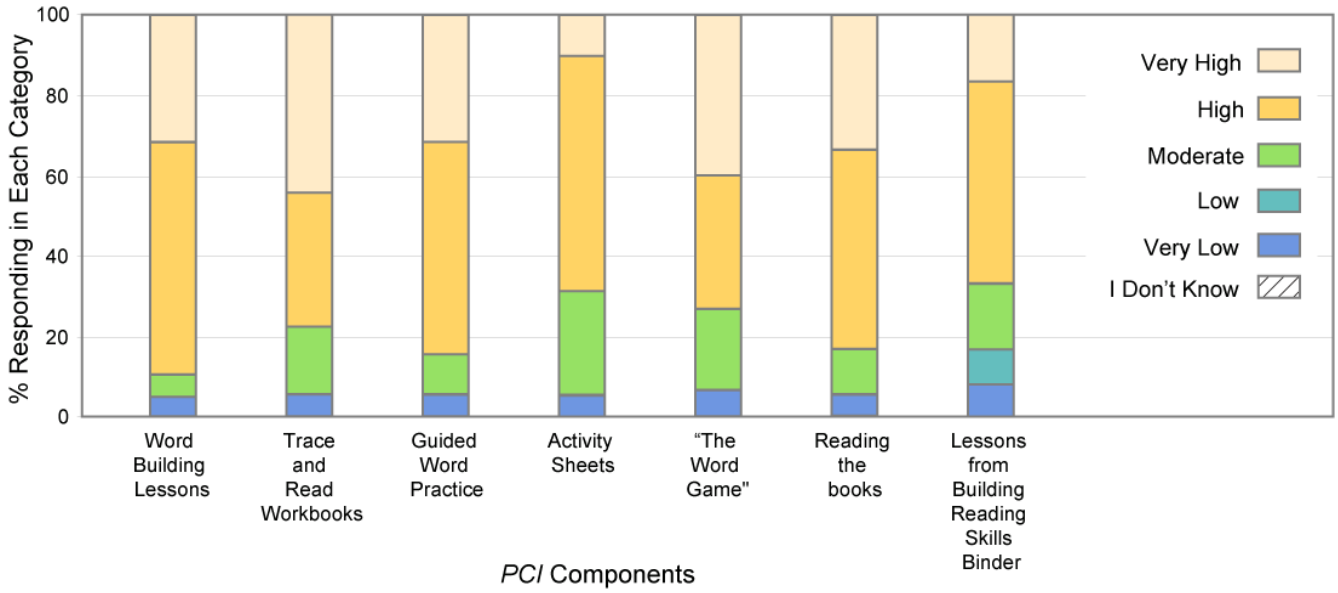

Figure 5. *Level One*: Levels of Student Engagement in Phase 2 with *PCI* Components[9]

[9] *n* = 20; *n* = 19 for word building lessons, guided word practice, and activity sheet; *n* = 18 for trace and read workbooks and reading the books; *n* = 15 for the word game; *n* = 12 for lessons from building reading skills binder.

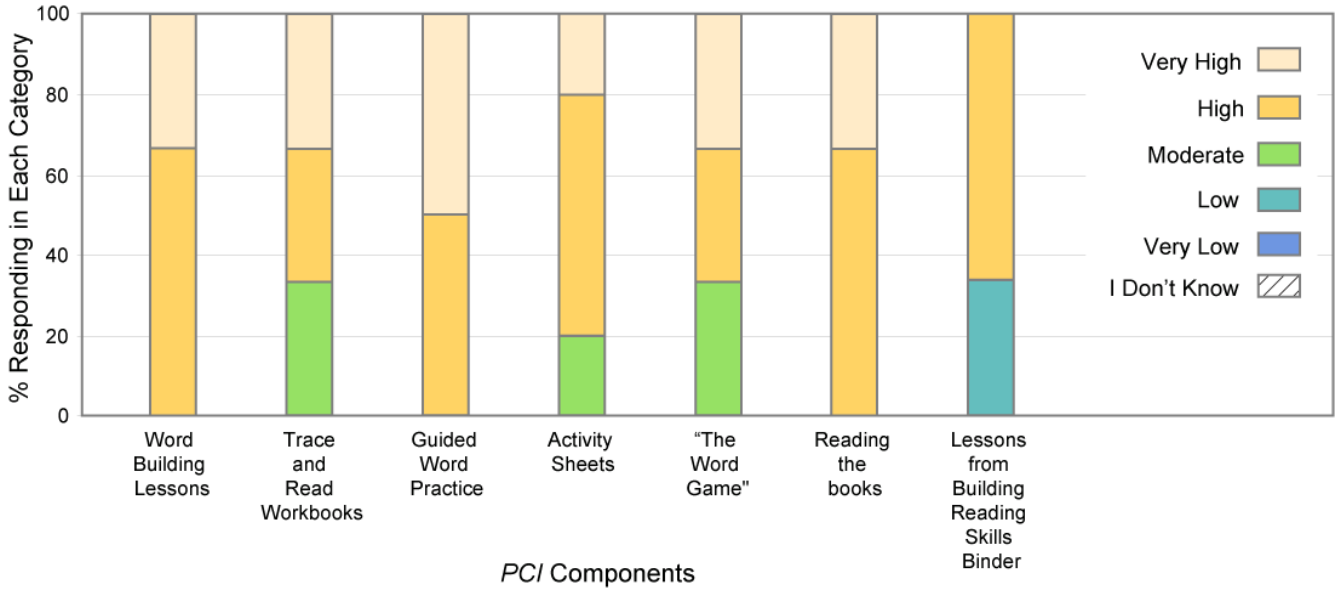

Figure 6. *Level Two*: Levels of Student Engagement in Phase 2 with *PCI* Components[10]

Although student engagement and teacher satisfaction were high, students weren't actually progressing through the program at the expected rate. By the end of the academic year, 83 students were on *Level One* of the program and six students were on *Level Two*. As displayed in Figure 7, almost half of the students that began with *Level One* ended the school year in between words 1-20 on *Level One*. Figure 8 shows that no student progressed beyond word 220 in *Level Two*.

[10] (*n* = 6); *n* = 6 for word building lessons; *n* = 6 for trace and read workbooks; *n* = 6 for guided word practice; *n* = 5 for activity sheet; *n* = 6 for the word game; *n* = 6 for reading the books; *n* = 3 for lessons from building reading skills binder.

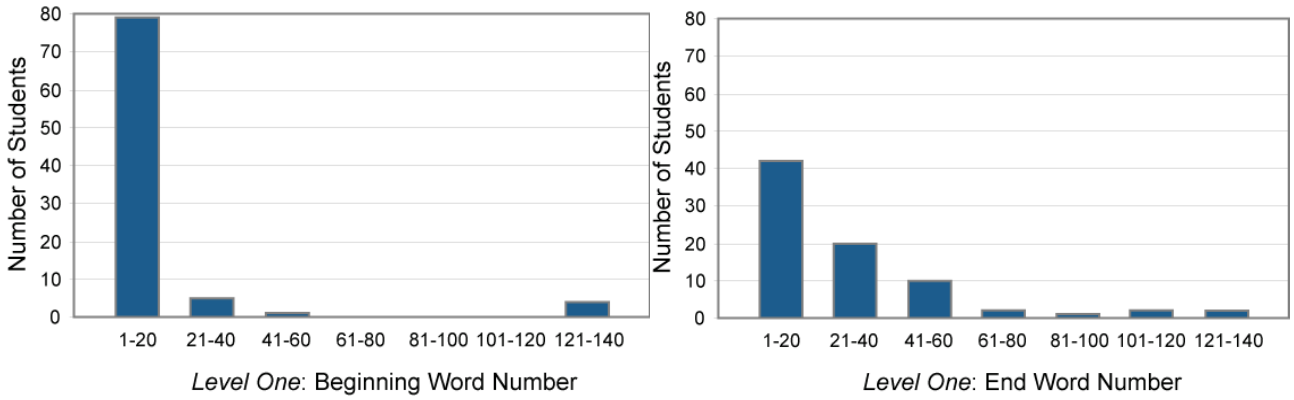

Figure 7. Student Progress: *Level One*

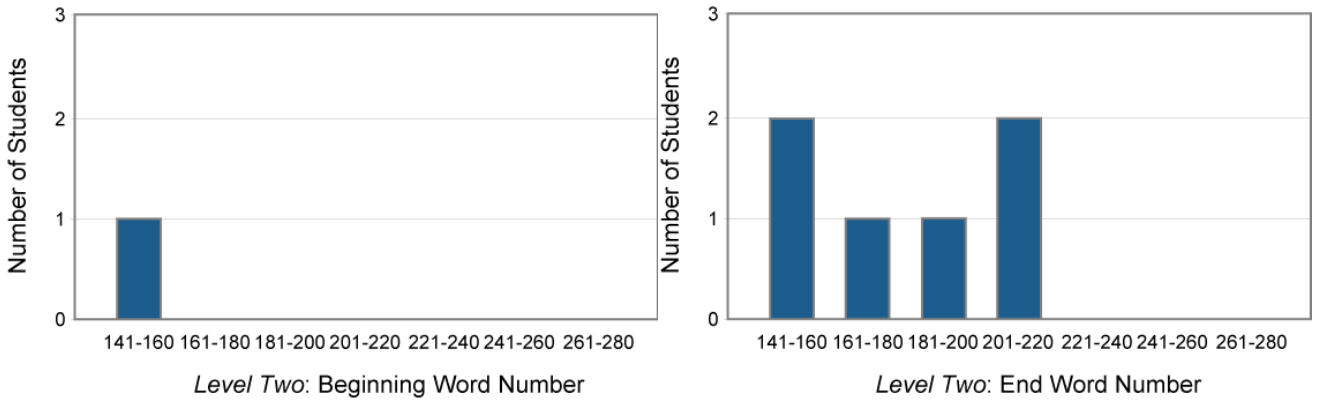

Figure 8. Student Progress: *Level Two*

Overall, comparison teachers reported more minutes of reading instruction than *PCI* teachers reported. In addition, 24% of the *PCI* teachers reported that they discontinued use of the program during FAA administration.

Qualitative data obtained from surveys, observations, and informal interviews in Phase 2 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 was 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

Phase 3 Implementation Results

In this section, we provide a description of the implementation of reading instruction in *PCI* and comparison classrooms during Phase 3. 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, and informal teacher interviews.

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.

Training

All first-year *PCI* teachers were offered training in the implementation of the *PCI* program and all five (100%) new teachers attended *PCI* training. All *PCI* teachers from Phases 1 and 2 who had been previously trained in the program were invited to attend an afternoon refresher session.

Figure 9 presents *PCI* teacher responses to survey questions regarding the effectiveness of the *PCI* training in preparing them to implement specific components of the *PCI* program. Among the teachers who had attended the full-day training and implemented the various components, all but two indicated that the training was moderately or very effective in preparing them to implement those components.

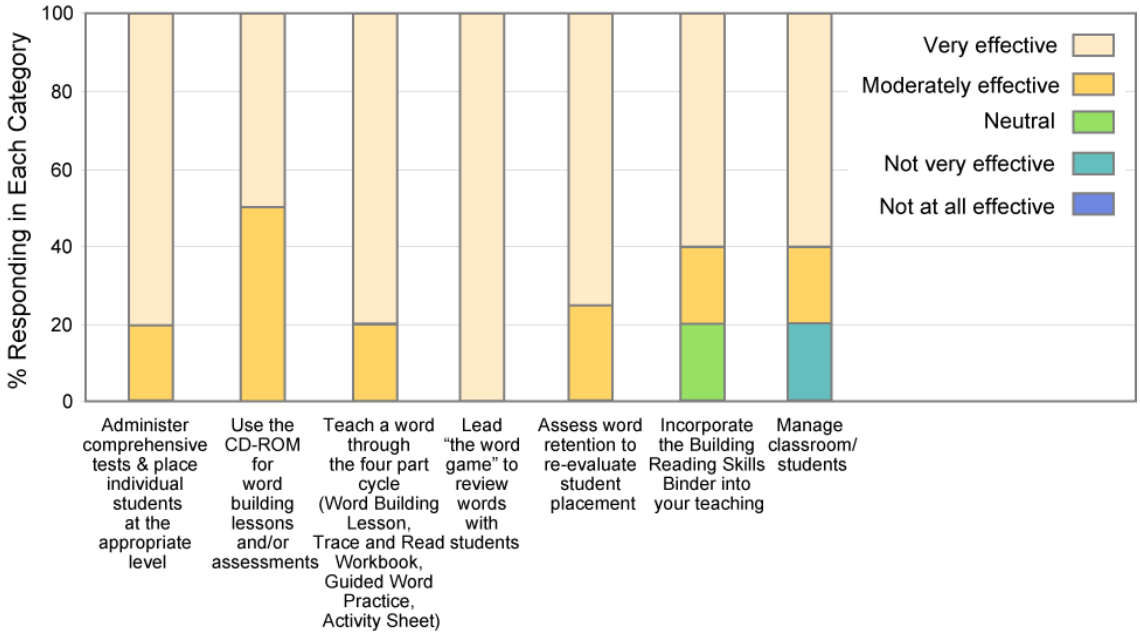

Figure 9. Effectiveness of Training by Task Area[11]

[11] *n* = 5 for administer comprehensive tests; *n* = 4 for use the CD-ROM; *n* = 5 for teach a word; *n* = 4 for lead "the Word Game"; *n* = 4 for assess word retention; *n* = 5 for incorporate the Building Reading Skills Binder; *n* = 5 for manage classroom/students.

Five of the seven (71%) comparison teachers reported receiving training for their current reading program, three of whom (60%) had been trained within the previous year. Of the five comparison teachers who received any training in their reading program, three reported it to be moderately to very effective, with the remaining teachers reporting a neutral opinion.

Availability of Materials

New *PCI* teachers who attended the training in November 2009 received the *PCI* materials at that time. *PCI* teachers from Phase 1 and 2 already had *PCI* materials by the beginning of the 2009-2010 school year. When surveyed in January 2010, 17 of 19 *PCI* teachers (89%) who responded to the survey reported that they had all the materials needed to fully implement the *PCI* program. One of the teachers who reported that they did not have all the materials needed was awaiting a *Level Two* kit; the other teacher needed additional Trace and Read notebooks as well as the reproducible binders. Five out of 7 comparison teachers (71%) reported having all the materials needed to fully implement their reading programs in January 2009.

Planning Time

Figure 10 displays the average number of hours teachers reported planning for reading instruction. Time spent planning for *PCI* teachers includes both *PCI* instruction and reading instruction with supplemental programs and materials. On average *PCI* teachers spent 3.7 hours per week planning for reading instruction, while comparison teachers spent 2.5 hours. The resulting *p* value of <.17 gives us limited confidence that this difference is not due to chance variation.

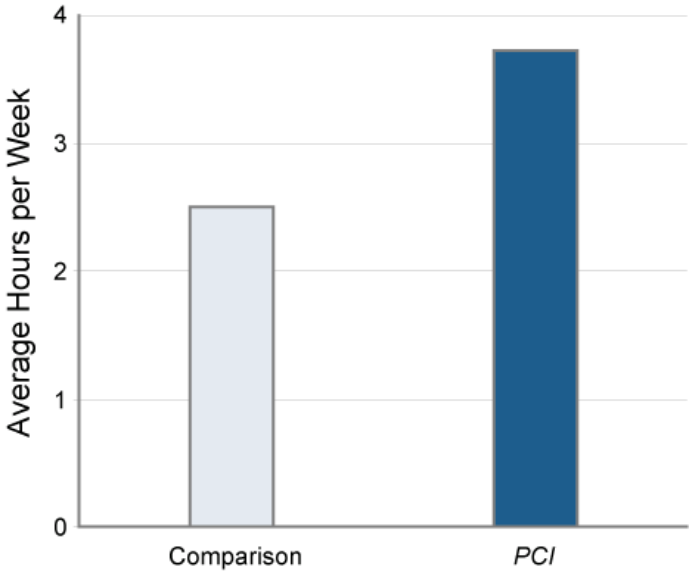

Figure 10. Average Planning Time[12]

[12] *n* = 7 for Comparison; *n* = 16 for *PCI*

Description of Implementation

Here we present our findings regarding classroom implementation during Phase 3. 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 classrooms.

Reading Materials Used

PCI Classrooms

Of the 15 Phase 3 teachers who had also taught *PCI* during Phases 1 and/or 2, nine (60%) began using *PCI* in August 2009 and continued through to the end of the year in May or June 2010. By December 2009, 19 of 19 (100%) *PCI* teachers had begun instruction with the *PCI* program.

In response to a survey question posed in March 2010, a majority of the *PCI* teachers reported supplementing reading instruction with other materials. Among the *PCI* teachers, 14 out of 18 (78%) who were teaching *Level One* and 4 out of 6 (67%) teaching *Level 2* reported using other materials to supplement reading instruction. By April 2010, 13 of 19 *PCI* teachers (68%) reported using supplemental materials.[13] A variety of materials were reportedly being used, including EdMark Functional Word series, The Letter People, Riverdeep, Unique Learning System, News-2-You, Sonday System, Scott Foresman Leveled Readers and a variety of teacher-made materials. During classroom observations, the researcher noted use of supplemental materials such as TumbleBook, News-2-You, Starfall, and Steck-Vaughn phonics.

Comparison Classrooms

Comparison teachers reported using a variety of materials for reading instruction such as Houghton Mifflin, Reading Milestones, Unique Learning System, News-2-You, Evan Moore, Sonday, and Lexia Learning. Teachers also noted use of picture cards, phonics exercises, and word lists. The researcher verified the use of News-2-You, Reading Milestones, and Evan Moore during classroom observations.

Teacher Satisfaction

At the end of the academic year, teachers in both assignment groups were asked about their level of satisfaction with their reading curriculum. As shown in Figure 11, 16 out of 17 *PCI Level One* teachers (94%) and 6 out of 6 *Level Two* teachers (100%) reported that they were very satisfied with the program, while only one comparison teacher reported that opinion of their reading program. None of the teachers in either the *PCI* or comparison group reported dissatisfaction with their reading program.

[13] One teacher that reported using supplemental materials during the March survey responded that she/he did not use supplemental materials in the April survey.

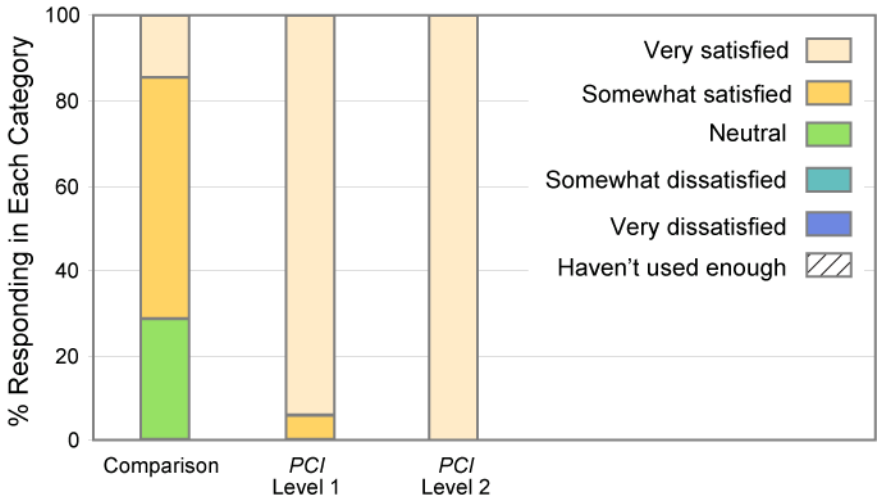

Figure 11. Teacher Level of Satisfaction with their Reading Program in Phase 3[14]

Figure 12 elaborates on how teachers rated the Building Reading Skills Binder (BRS) and the reproducible story books (RSB), which are optional or supplemental components of the *PCI* program. Other than teachers who had not used each piece of the program enough to form an opinion and the two teachers who selected "neither satisfied nor dissatisfied" for the Level 1 BRS, all other respondents indicated that they were satisfied or very satisfied with these aspects of the program.

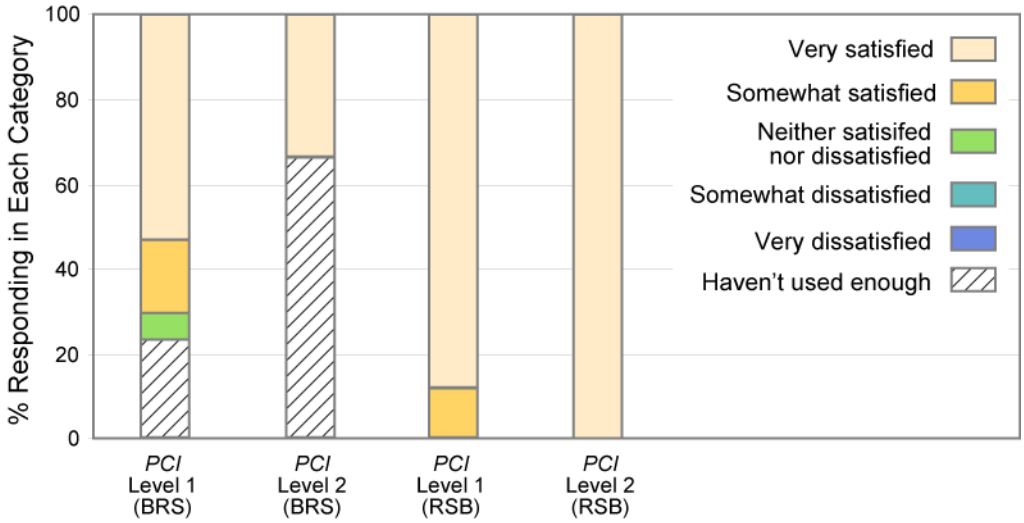

Figure 12. Teacher Satisfaction with Optional/Supplemental *PCI* Components in Phase 3[15]

[14] *n* = 7 for Comparison; *n* = 17 for *PCI Level One*; *n* = 6 for *PCI Level Two*

Researchers also asked teachers in both conditions whether they would recommend their reading program to other teachers. All teachers using Level One and all teachers using Level Two who provided valid responses (17 and 6 respectively) reported that they would recommend the program. These numbers are slightly higher than in Phase 2. In contrast, only four of the seven comparison teachers (57%) who provided valid responses said they would recommend their current program to teachers of this population. The remaining teachers in the comparison group responded that they did not know whether they would recommend their reading program.

Table 23. Would you recommend this reading program to other teachers?

	Yes	No	I don't know
Comparison (*n* = 7)	4 (57.1%)	0 (0.0%)	3 (42.9%)
***PCI Level One* (*n* =17)**	17 (100%)	0 (0.0%)	0 (0.0%)
***PCI Level Two* (*n* = 6)**	6 (100%)	0 (0.0%)	0 (0.0%)

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

- repetition and review within the program—structure and routine
- students engagement and enjoyment of reading books
- program is flexible and can address the needs of a variety of students
- reproducible worksheets and books to take home and share
- individual components such as books, picture cards, computer work, games, and word cards

The primary difficulty reported by *PCI* teachers was classroom management, and finding the time for individualized instruction, since much of the program is administered one-on-one. This echoes teacher comments from both Phase1 and Phase 2 of the study. Other teachers commented on difficulty with obtaining additional or replacement materials. Several teachers noted that some of their students had more difficulty with the program than others.

Within the comparison group, teachers commented both positively and negatively about the variety of materials used in their classrooms. Some teachers appreciated the variety in that it

[15] *n* = 17 BRS for *PCI Level One*; *n* = 17 BRS for RSB for *PCI Level One*; *n* = 6 for BRS for *PCI Level Two*; *n* = 5 for RSB for *PCI Level Two*

allowed them to differentiate instruction for meeting the needs of individual students, while others enjoyed more structure within their reading programs. Teachers complained in large numbers about the lack of a comprehensive program for students of this population, claiming their current materials aren't particularly engaging for their students.

PCI Reading Program Levels of Implementation

Lesson Cycle

Twice over the course of the school year – once at the beginning of program implementation and once toward the end –teachers were surveyed about the extent to which they followed the prescribed lesson cycle. As in Phases 1 and 2, the Word Game is one of the commonly skipped steps. This is not surprising when considering the low levels of reported student engagement with this task. The four components of Step 1, in addition to the Read a Book step, were adhered to with the most fidelity, with over 70% of teachers reporting that they always teach these steps. Teachers using both *Level One* and *Level Two* reported that most of the instruction was conducted in a one-on-one manner and that instruction was divided between the teacher and other adults in the classroom. Teachers generally taught each step more than other adults in the classroom, although other adults played a significant part of *PCI* instruction. See the Appendix for detailed teacher responses.

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. In February 2010, teachers were asked how students usually perform on the posttest. Table 24 shows that, for *Level One* and *Level Two,* more than half of the teachers reported that their students master a majority of the words. Two *Level One* teachers who selected "Other" stated that some of their students have problems retaining the words in general and one other teacher reported that they had a student who couldn't pass the first lesson and was using the Building Reading Skills binder instead. The *Level Two* teacher who selected "other" did not explain.

Table 24. Student Performance on Assessment in Lesson Cycle

	I have students who master the majority of the words	I have students who master the most recent five words, but have difficulty retaining words taught in previous lessons	I have students who have difficulty with both old and new words	Other
***Level One* (*n* = 18)**	13 (72.2%)	6 (33.3%)	6 (33.3%)	3 (16.7%)
***Level Two* (*n* = 7)**	6 (85.7%)	0 (0.0%)	0 (0.0%)	1 (14.3%)

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

A February survey also asked teachers what they usually do if a student misses a word on the posttest. 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." Table 25 shows that 14 out of the 18 *Level One* teachers

(79%) and 5 out of 7 *Level Two* teachers (71%) adhered to the Teacher's Guide and would re-teach the word lesson if a student missed a word on the posttest. No teacher in either group reported that they skipped the review and moved on to keep pace with other students. The teachers who reported "Other" made several different accommodations. One teacher re-taught the word using methods such as displaying the word around the room, making a "pretty card" for the word, or making a rainbow word. Another teacher reported that they re-administered the post-test if a student did not pass it, and yet another teacher said they re-taught the missed word separately but continued with the other words.

Table 25. 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)**	14 (78.8%)	6 (33.3%)	0 (0.0%)	4 (22.2%)
***Level Two* (*n* =7)**	5 (71.4%)	3 (42.9%)	0 (0.0%)	1 (14.3%)

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

Bonus Materials

The Activity Sheets are part of the lesson cycle and, because they are reproducible, teachers have the option of assigning them as homework. As of May 2010, of the 17 *Level One* teachers who responded, 13 (76%) reported using the resource for in-class exercises, nine (53%) sent Activity Sheets as homework, and two teachers (12%) reported never having used the resource at all. Of the six *Level Two* teachers who responded, all six reported using the resources for in-class exercises, and two (33%) 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 17 *Level One* teachers who responded, no teachers reported using only the CD-ROM for these tasks, ten (59%) reported using only the print materials, and six (35%) used both the CD-ROM and print materials at some point. Of the six *Level Two* teachers, one (17%) reported using only the CD-ROM for these tasks, three (50%) reported using only print materials, and two teachers (33%) used 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, including help with phonics. By May 2010, ten out of 17 teachers (59%) had utilized this resource. While the binder is optional in *Level One,* it is expected to be used in *Level Two* to prepare students for *Level Three*. However, in May, of the six teachers on *Level Two*, only one reported using the binder with their students. This is similar to what was reported in Phase 2.

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

Continued Use

As in Phase 1 and Phase 2, *PCI* teachers in Phase 3 expressed satisfaction with 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. Sixteen out of the 17 *Level One* teachers who responded (94%) and 5 out of the 6 *Level Two* teachers (83%) planned to continue to use the program. One teacher said they planned to discontinue using the program. (See Table 26)

Table 26. 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* = 17)**	10 (58.8%)	6 (35.3%)	0 (0.0%)	1 (5.9%)	0 (0.0%)
***Level Two* (*n* = 6)**	2 (33.3%)	2 (33.3%)	1 (16.7%)	1 (16.7%)	0 (0.0%)

Summary

As in Phases 1 and 2, *PCI* teachers continued to show enthusiasm for and satisfaction with the program. Teachers generally followed the lesson cycle and used the additional supplemental materials; however, *Level Two* teachers were still not utilizing the Building Reading Skills Binder as intended by the publisher. Nearly all teachers said they would continue using the program.

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 classrooms during a specified week. On average comparison teachers reported that students received 192 minutes per week and *PCI* teachers reported 206 minutes per week. In the *PCI* group, this number reflects both *PCI* instruction and instruction with other/supplemental reading materials.

Figure 13 shows the average minutes of reading instruction students received in their classrooms over the course of the academic year surveyed. In contrast to Phase 2, comparison teachers in Phase 3 reported fewer minutes of reading instruction than the *PCI* group. However, with $p<.80$, we have no confidence that this difference is not simply due to chance. Because of the small sample size in Phase 3, we were not able to investigate any correlation between minutes of instruction and student outcomes.

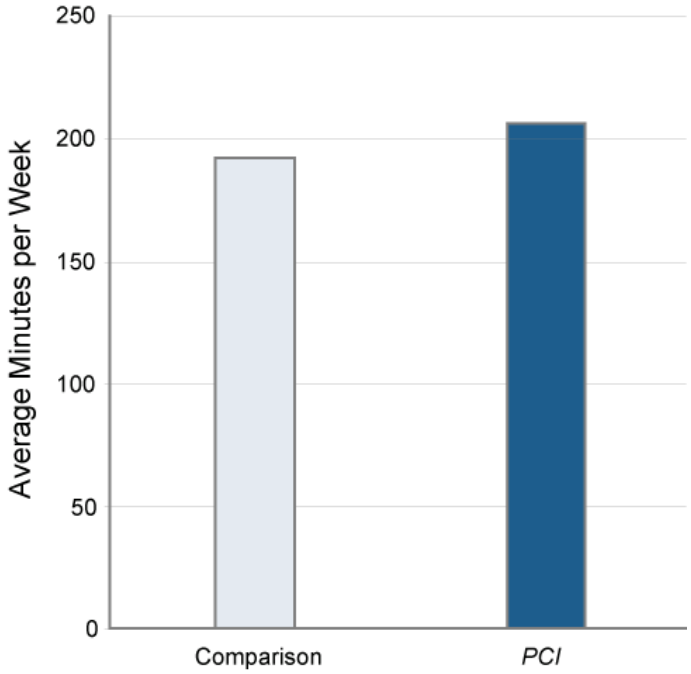

Figure 13. Average Weekly Minutes of Reading Instruction[17]

Surveys also asked both assignment groups whether they had stopped using or supplemented their current reading program in order to focus on FAA test preparation. As displayed in Table 27, not one comparison teacher reported discontinuing the use of their reading program in preparation for the FAA. In fact, 71% reported no impact on instruction. However 4 out of 19 *PCI* teachers (21%) reported that they stopped *PCI* instruction to prepare for the FAA.

Table 27. What is the impact, if any, of the FAA (Florida Alternate Assessment) on your *PCI* instruction?

	There is no impact	I spend less time using my current reading program	I stopped using my current reading program	Other
Comparison (*n* = 7)	5 (71.4%)	2 (28.6%)	0 (0.0%)	0 (0.0%)
***PCI* (*n* = 19)**	3 (15.8%)	6 (31.6%)	4 (21.1%)	6 (31.6%)

[17] Analytic sample sizes varied across surveys.

Student Engagement

Researchers asked teachers in both groups to rate the average level of student engagement with their reading programs. As in previous phases, teachers were instructed to consider students as fully engaged if they displayed consistent on-task behavior. Table 28 displays that 11 out of 17 *Level One* teachers (65%) and 7 of eight (88 %) *Level Two* teachers reported that their students were highly or very highly engaged with the program. Five out of 7 comparison teachers (71%) reported moderate to high levels of engagement, which is higher than in previous phases.

Table 28. Level of Student Engagement

	Very low	Low	Moderate	High	Very high	I don't know
Comparison (*n* = 7)	0 (0.0%)	0 (0.0%)	2 (28.6%)	5 (71.4%)	0 (0.0%)	0 (0.0%)
***PCI Level One* (*n* =17)**	1 (5.9%)	2 (11.8%)	3 (17.7%)	6 (35.3%)	5 (29.4%)	0 (0.0%)
***PCI Level Two* (*n* = 8)**	0 (0.0%)	0 (0.0%)	1 (12.5%)	2 (25.0%)	5 (62.5%)	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 14 and Figure 15, 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.

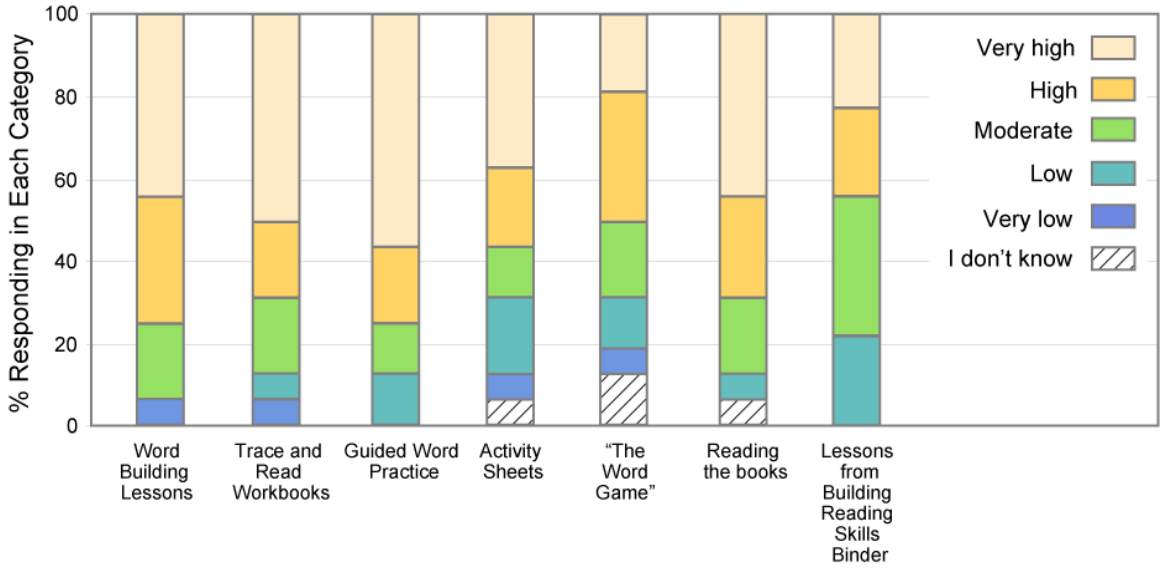

Figure 14. *Level One*: Levels of Student Engagement with *PCI* Components in Phase 3[18]

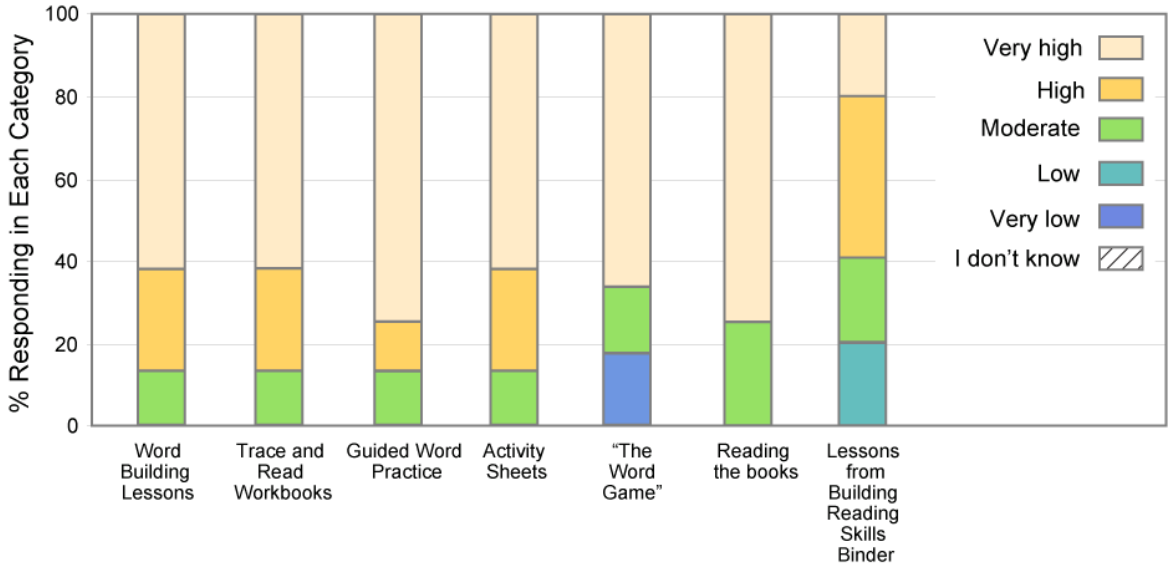

Figure 15. *Level Two*: Levels of Student Engagement with *PCI* Components in Phase 3[19]

[18] *n* = 16 for all components except for lessons from Building Reading Skills Binder. *n* = 9 for lessons from Building Reading Skills Binder due to the fact that eight teachers reported "N/A."

Figure 14 shows that the word building lessons, guided word practice, and reading books were ranked very high in terms of student engagement, while several teachers were unable to rank student engagement in The Word Game and lessons from the Building Reading Skills binder, possibly because they did not use this tool enough.

Summary

As in Phase 1 and Phase 2, *PCI* teachers continued to report a high level of student engagement with *PCI* although teachers in the control condition reported high levels of student engagement, as well. In addition, *PCI* teachers expressed high levels of satisfaction with the program and the intent to continue using the program after the research study has concluded. In contrast to Phase 2, comparison teachers reported fewer minutes of reading instruction than *PCI* teachers reported, although that difference is not statistically significant.

Student-level Impact Results

In this section we first summarize the student-level impact results from Phases 1 and 2 and then discuss the findings from Phase 3 with regard to student progress through the program. Since we deemed the sample size too small to carry out quasi-experimental or extra-experimental analyses of the impact of the *PCI* program on the Sight word or Phonological assessments, we did not conduct those analyses in Phase 3.

Summary of Student-Level Impact Results from Phases 1 and 2

The following tables summarize the student impact results of Phase 1 and Phase 2 of this study. Table 29 summarizes the main impact results and addresses potential for bias due to selection, changing condition, and attrition.

---

[19] *n* = 8 for all components except for the Word Game (*n* = 6) and lessons from the Building Reading Skills binder (*n* = 5). For the Word Game, *n* = 6 because one teacher did not answer to the question, and one teacher reported "N/A." For lessons from the Building Reading Skills binder, *n* = 5 as three teachers reported "N/A."

Table 29. Sight Word Reading Performance

Impact	Method	Estimate[a]	Bias due to initial selection into conditions	Bias due to secular changes affecting all groups	Bias due to attrition
1-year	Experimental	3.17 ($p < .05$)	Ruled out	Ruled out	Passes checks, but possible
2-year	Quasi-experimental	6.12 ($p = .06$)	Passes checks, but possible	Ruled out	Passes checks, but possible
2-year	Extra-experimental	5.81 ($p = .02$)	Ruled out	Ruled out[b]	Passes checks, but possible

[a] In units of the outcome scale

[b] bias may still be present if the program continues to be under development so that the measured two-year impact does not reflect what the two-year impact would be with the new iteration of the program.

In Phase 1, despite attrition among students and teachers, the experiment was able to detect an impact on sight word learning equivalent to a 21 percentile point difference between students in the *PCI* program and those in the comparison group. There was a sizable number of students who were unable to name any words on the pretest. Researchers believed that the composition of this group of students differed fundamentally from the rest of the students in the study; therefore, we conducted separate statistical tests for those scoring zero on the pretest and those scoring above zero. Within each group, we found a significant impact for the *PCI* program. Both the unadjusted analysis and the analysis where we adjusted for the effects of covariates show high effect sizes of .55 and .59 standard deviation units, respectively, with small *p* values.

In Phase 2, with both the quasi-experimental and extra-experimental approaches to estimating the two-year impact of *PCI*, we found that students in the *PCI* classrooms outperformed students who were not exposed to the program on the Sight word assessment. The quasi-experimental analysis matched students who had received *PCI* to students who had never received *PCI* and measured the regression-adjusted differences between them in performance. The .06 *p* value gives us some confidence in the difference found. The adjusted effect size of .89,is equivalent to a difference of 31 percentile points. Using the extra-experimental approach we estimated the difference in performance between students who were members of the originally randomized groups after the first and second years, and combined these estimated differences to infer what the performance level would have been had a true control group been maintained throughout the first two years of the study. The difference found using the extra-experimental approach (an adjusted effect size of .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* is larger after two years than it is 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. The table also describes the possibilities for bias due to selection, secular changes affecting performance, or attrition.

Moderating Effects

Table 30 summarizes the moderating effects of certain variables.

Table 30. 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[a]	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 samples to run this analysis.

In examining potential moderating variables in Phase 1, we found no moderating effect of pretest or grade level on the impact of *PCI* on student outcomes. However, we have some confidence that the Phonological assessment had a small moderating effect, i.e., students starting with greater phonological skills benefit more from *PCI* than students starting with lower phonological skills. The small number of teachers with fewer than four years experience teaching Special Education prevented us from examining moderating effects of teacher experience.

When examining potential moderator effects using the quasi-experimental approach in Phase 2, we found that the Sight word pre-assessment did not moderate the impact of *PCI* on Sight word post-assessment scores. 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. Due to the sample size and imbalance between the two groups on the Phonological pre-assessment, these were the only potential moderators we were able to examine with the quasi-experimental approach. None of the estimates of moderating effects of the two year impact based on the extra-experimental approach reached significance.

Word Gain Results Phases 1-3

Due to the small number of students remaining in the *PCI* group (there were only six students remaining who had received the program for three years), we were not able to conduct analyses of the three year impact at the end of Phase 3[20]. Instead, we focused on students in the *PCI* group and their progress through *Levels One* and *Two* of the program. Initially, PCI expected students to be able to

[20] Another possibility would have been to run quasi-experimental two-year impact analyses in Phase 3. However, the pool of comparison cases was too small to go forward with this, and the results from Phase 2 would be more definitive regardless.

master one level (140 sight words) in one school year. This rate of progress assumes students are receiving the recommended 45 minutes of *PCI* instruction per student per day. In Phase 1 and again in Phase 2 we learned through teacher-reported minutes of instruction and additional teacher survey comments that teachers were not able to devote the recommended 45 minutes daily to *PCI* instruction and that students were not advancing as quickly through the program as the publisher had intended. Therefore, in Phase 3 we investigated the minutes of instruction alongside student word progress and again found evidence that student progress through the levels was slower than expected.

In Phase 3, we found that teachers were spending more time using *PCI* than in the previous years. In fact, *PCI* students were receiving an average of 40 minutes a day of *PCI* instruction, only five minutes short of the publisher's recommendation.

Also during Phase 3 we probed for reasons for the slower than expected progress. Specifically, researchers surveyed teachers about whether specific subgroups of students were progressing more readily than others. However, we did not find any patterns to suggest that specific subgroups of students progressed more rapidly than others.

Researchers also investigated the number of words students were able to learn over the course of one year and two years using *PCI*. During Phases 2 and 3 teachers reported each student's beginning word and end word (i.e. the word the student was working on at the time of the end-of-year questionnaire) for that particular school year. This was accomplished in Phases 2 and 3 of the study. However, in Phase 1 of the study, researchers only obtained the end word for a sample of students. Because we did not collect start and end word data for all students in Phase 1, the data in this section represent only the years for which we have data, which are the 2008-2009 and 2009-2010 school years.

To determine how many words students learned, or the "word gain," we subtract the word number the student started with at the beginning of the year from the word number the student was on at the end of the school year. Figure 16 displays our exploration of how many words students learned within one year of *PCI* instruction. This graph includes and highlights separately students who received only one year of *PCI* instruction during the study and students who received a second year of *PCI* instruction (although the figure shows their gain for only their first year). Because the data are skewed, we report the median number of words gained as opposed to the average. Again, because we do not have start and end words for Phase 1, the first year of instruction would be either the 2008-2009 or 2009-2010 school year.

The majority of students don't progress beyond 10 words, similar to what we observed in Phase 2 of the study. The median number of words gained for students with only one year of exposure to *PCI* was 15. The median number of words gained in the first year for students with two years of exposure to the program was 14. Overall, the median number of words gained after the first year of *PCI* was 14.5 (as compared to the expected 140). Only 4 out of 104 students reached the expected gains in a year's instruction time.

While the number of words mastered by the *PCI* students was lower than expected, many teachers commented through the online surveys that they felt their students were making great progress. Some also commented that this perceived progress was not detected by the assessments. So the teachers' generally positive view of the program may have resulted from student gains in other areas such as attitudes toward reading.

The teachers' view of student progress may also have been due to expectations in relation to the severity of the student's disability rather than their progress relative to the progam's expecations. We do know that there is great variation in student disabilities within this project but we are unable to obtain specific classifications from the district. Therefore, the students PCI envisions using the program and the students who are using the program in this study may be somewhat different.

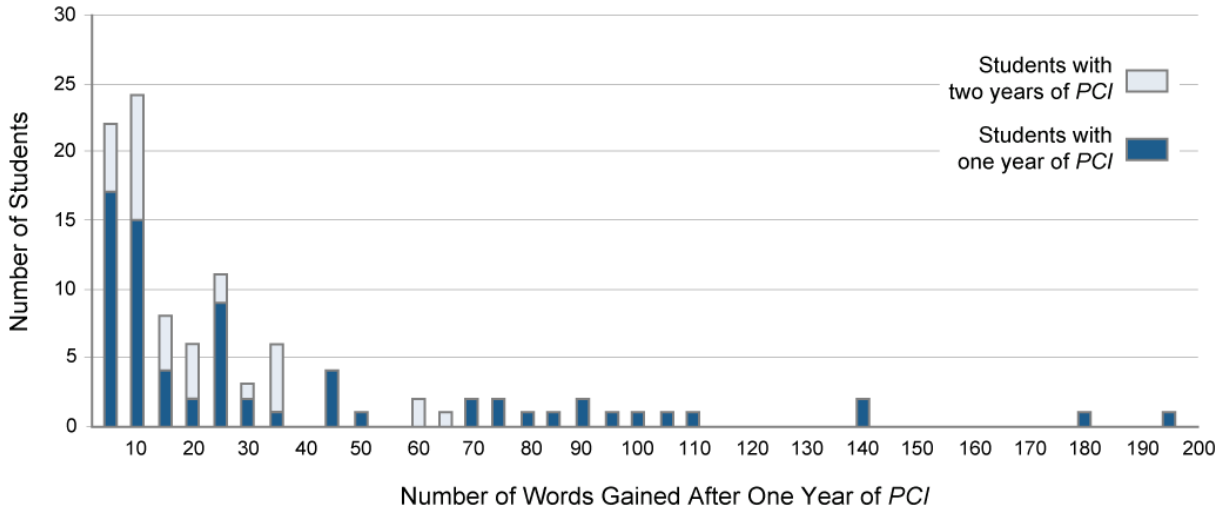

Figure 16. Words Gained in One Year

In Figure 17, we display our exploration of the number of words gained over two years of exposure to the *PCI Reading Program*. That is, we are looking at the end of second year word minus the start of the first year word. In this instance we have fewer students because we included only those students who received two consecutive years of instruction in the *PCI* program (*n* = 27).

The median students gained 33 words over the two-year span, which is more than the number of words mastered by those with only one year of exposure to *PCI*. Therefore students continue to progress and master more words with additional years of exposure to the program. In addition, the rate of progress increases in the second year. The median number of words gained during the second year of exposure to *PCI* is 16.5 for the students represented in Figure 17.

We also see a more normal distribution in the case of students with two years of exposure. Fewer students remain stuck with only 5-10 words gained over the course of one school year.

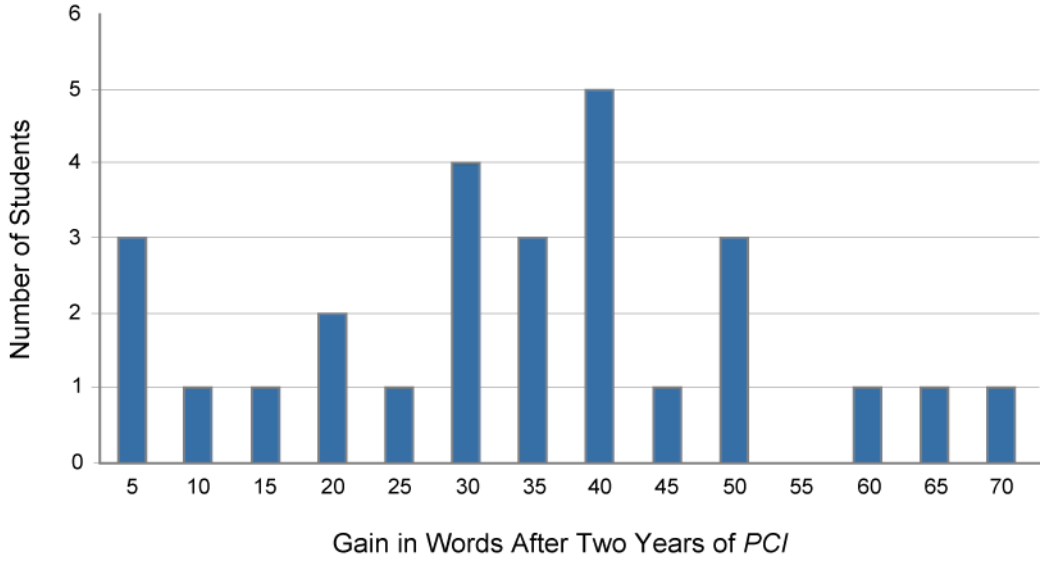

Figure 17. Word Gain After Two Years

The following two figures are included to elaborate on above Figure 16 and Figure 17 as an additional narrative. In these two graphs we see the progression of students between the different years of exposure to the program. Figure 18 shows us the start and end word for students in their first year of *PCI.* This graph includes all students with at least one year of exposure but includes progress in both *Level One* and *Level Two*. Again, because of the volume of students and style of graph, it is difficult to discern any noticeable trend. What we can observe is that students generally made positive progress, although much of that progress is bunched toward the lower end of the graph, between words zero and fifty.

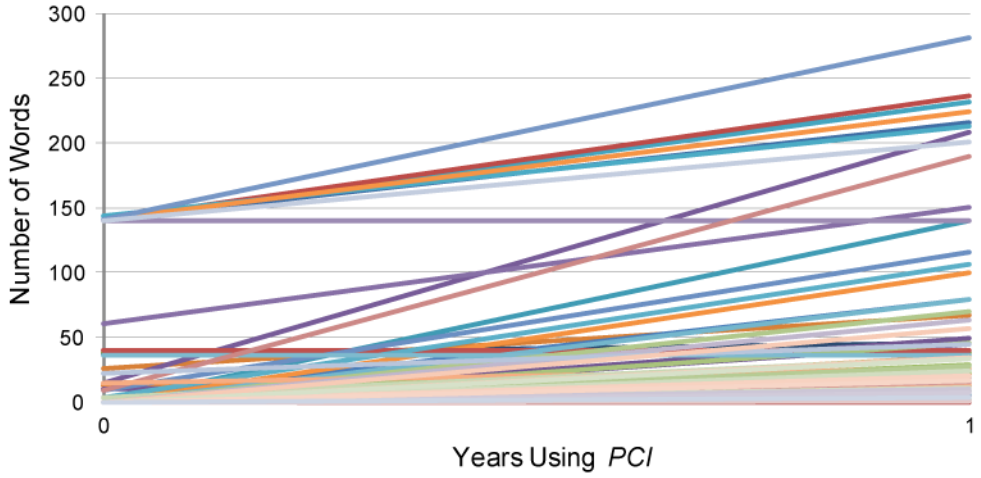

Figure 18. Number of Words Mastered after One Year with *PCI*

Figure 19 provides a visual demonstration of how the means for students with two years of exposure can be lower than for students with one year of exposure. Here we have fewer students than in Figure 18 and are able to obtain a clearer view of individual student progress. We see that most students have a steady positive slope in their first year of receiving *PCI* instruction. For some students this pace continues, for others it accelerates, and for others progress plateaus or even decreases.

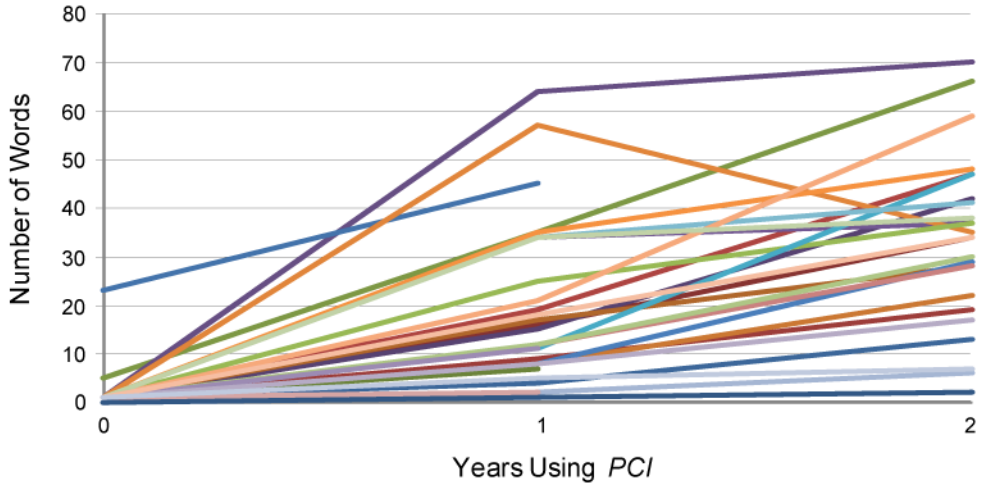

Figure 19. Number of Words Mastered after Two Years with *PCI*

In conclusion, in Phase 3 analysts were able to measure the number of words students actually mastered through the course of the program (end word minus start word). We found that within one year of exposure students make steady uphill progress and the median students gain 14.5 words as opposed to the 140 words as initially predicted by the publisher. The rate at which students learn words actually increases during a second year of exposure to the program. Therefore students appear to benefit greatly from continued exposure to the program.

Discussion

This report contains the findings from a three-year longitudinal study on the efficacy of the *PCI Reading Program* as implemented in two Florida school districts. 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 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. 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 progressed through the program. 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.

The Phase 1 research was a randomized control trial with teacher-level randomization. In Phase 2, students in experimental control classes joined the *PCI* group, and a new set of comparisons students were recruited. 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 the comparison students who had not been exposed to *PCI* and 2) using an extra-experimental approach based on the differences in performance at the end of Phases 1 and 2 for students who were members of the originally randomized classes. The Phase 2 research therefore 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 also 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.

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 .89 and .98, respectively) with small *p* values (.06 and .02, respectively). These were consistent with the significant effects found in the initial RCT phase (.55 and .59 standard deviations for the unadjusted and adjusted analyses) and consistent with the expectation that two years of growth would be greater than 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 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).

Phase 3 of the study corroborated most of what we had learned, with regard to program satisfaction, from Phases 1 and 2. Teachers continued to feel well-prepared for teaching *PCI* after the initial training and general conditions for implementing the program, including access to all needed materials, were very good. Throughout all three phases of the study teachers consistently reported high levels of satisfaction with the program, intent to continue teaching *PCI* at the conclusion of the research study, and an endorsement of the program to other reading teachers of similar populations. In addition, teachers reported high levels of student engagement and perceived enjoyment with *PCI*. Teachers across all three phases continued to supplement *PCI* instruction with a variety of other materials. Teachers also consistently commented on the struggle to find time for the individualized instruction required for ideal implementation of the program.

In spite of the demonstrated learning of sight words, it was clear that for many students, learning to recognize words was limited. 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.

This study required a complex design that was unfortunately not sustainable within our limited context for the intended duration. Though our research certainly could have benefitted from a larger sample, we are satisfied to be able to publish research on a reading program for a population of students that is in great need of supportive literature. Here, we are able to confidently present evidence of a reading program for students with supported level disabilities, that does teach students to read words. Furthermore, we find evidence that such students become engaged in the process and enjoy doing so. With a larger sample size, continuing research on this reading program could investigate how students attain and retain sight words over time and further explore any patterns regarding student movement through the program. This population of students, in addition to the reading community as a whole, would benefit from further research into this and similar reading programs.

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 (2011, January).Retrieved on January 12, 2011 from http://www.fldoe.org/eias/eiaspubs/#student

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

Table A1. *Level One*: Which steps do you usually complete during each lesson cycle?

	Always	Sometimes	Never	NA
Step 1a. Learn a Word: Word Building Lesson	88.3%	2.8%	0.0%	2.8%
Step 1b. Learn a Word: Trace and Read Workbook	79.5%	11.7%	0.0%	2.8%
Step 1c. Learn a Word: Guided Word Practice	76.1%	15.0%	0.0%	2.8%
Step 1d. Learn a Word: Activity Sheets	55.0%	33.0%	2.8%	2.8%
Step 2. Repeat steps 1a-1d to learn four new words	51.7%	30.1%	2.8%	6.1%
Step 3. Review the words with "The Word Game"	25.0%	51.1%	12.2%	2.8%
Step 4. Assess Word Retention	49.5%	39.5%	0.0%	2.8%
Step 5. Read a Book	70.0%	24.5%	0.0%	2.8%

Table A2. *Level Two*: Which steps do you usually complete during each lesson cycle?

	Always	Sometimes	Never	NA
Step 1a. Learn a Word: Word Building Lesson	83.3%	8.3%	0.0%	0.0%
Step 1b. Learn a Word: Trace and Read Workbook	91.7%	0.0%	0.0%	0.0%
Step 1c. Learn a Word: Guided Word Practice	73.3%	18.3%	0.0%	0.0%
Step 1d. Learn a Word: Activity Sheets	71.7%	20.0%	0.0%	0.0%
Step 2. Repeat steps 1a-1d to learn four new words	55.0%	20.0%	16.7%	0.0%
Step 3. Review the words with "The Word Game"	40.0%	43.3%	8.3%	0.0%
Step 4. Assess Word Retention	48.3%	48.3%	0.0%	0.0%
Step 5. Read a Book	75.0%	16.7%	0.0%	0.0%

Table A3. *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	79.4%	23.9%	0.0%	2.8%
Step 1b. Learn a Word: Trace and Read Workbook	52.2%	17.8%	40.0%	5.6%
Step 1c. Learn a Word: Guided Word Practice	82.8%	17.2%	0.0%	2.8%
Step 1d. Learn a Word: Activity Sheets	24.4%	40.0%	61.1%	5.6%
Step 2. Repeat steps 1a-1d to learn four new words	82.8%	23.9%	3.3%	5.6%
Step 3. Review the words with "The Word Game"	39.4%	36.1%	3.3%	15.0%
Step 4. Assess Word Retention	85.6%	17.8%	0.0%	2.8%
Step 5. Read a Book	82.2%	30.6%	11.7%	2.8%

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

Table A4. *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	91.7%	8.3%	8.3%	0.0%
Step 1b. Learn a Word: Trace and Read Workbook	63.3%	0.0%	46.7%	0.0%
Step 1c. Learn a Word: Guided Word Practice	81.7%	28.3%	18.3%	0.0%
Step 1d. Learn a Word: Activity Sheets	35.0%	28.3%	65.0%	0.0%
Step 2. Repeat steps 1a-1d to learn four new words	75.0%	10.0%	8.3%	16.7%
Step 3. Review the words with "The Word Game"	75.0%	18.3%	8.3%	8.3%
Step 4. Assess Word Retention	91.7%	0.0%	0.0%	0.0%
Step 5. Read a Book	91.7%	0.0%	55.0%	0.0%

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

Table A5. *Level One*: Who is this step usually taught by?

	Teacher	Other adult	NA
Step 1a. Learn a Word: Word Building Lesson	78.9%	30.0%	2.8%
Step 1b. Learn a Word: Trace and Read Workbook	70.6%	45.0%	5.6%
Step 1c. Learn a Word: Guided Word Practice	82.2%	38.9%	2.8%
Step 1d. Learn a Word: Activity Sheets	77.2%	45.6%	8.3%
Step 2. Repeat steps 1a-1d to learn four new words	82.2%	36.1%	5.6%
Step 3. Review the words with "The Word Game"	67.2%	38.9%	18.3%
Step 4. Assess Word Retention	82.2%	30.6%	2.8%
Step 5. Read a Book	85.6%	48.3%	2.8%

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

Table A6. *Level Two*: Who is this step usually taught by?

	Teacher	Other adult	NA
Step 1a. Learn a Word: Word Building Lesson	65.0%	63.3%	0.0%
Step 1b. Learn a Word: Trace and Read Workbook	46.7%	63.3%	0.0%
Step 1c. Learn a Word: Guided Word Practice	46.7%	63.3%	0.0%
Step 1d. Learn a Word: Activity Sheets	36.7%	55.0%	18.3%
Step 2. Repeat steps 1a-1d to learn four new words	46.7%	46.7%	16.7%
Step 3. Review the words with "The Word Game"	55.0%	55.0%	8.3%
Step 4. Assess Word Retention	55.0%	63.3%	100.0%
Step 5. Read a Book	65.0%	63.3%	10.0%

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