



---

# **Results From a Three-Year i3 Impact Evaluation of the Children's Literacy Initiative (CLI)**

**Implementation and Impact Findings of  
an Intensive Professional Development  
and Coaching Program**

JUNE 2015

**Results From a Three-Year i3 Impact  
Evaluation of the Children’s Literacy  
Initiative (CLI):**

**Implementation and Impact Findings of  
an Intensive Professional Development  
and Coaching Program**

**June 2015**

**Julia Parkinson**

**Terry Salinger**

**John Meakin**

**Deeza-Mae Smith**

*American Institutes for Research*

# Contents

- Overview ..... 1
- The Children’s Literacy Initiative Program ..... 2
- i3 Evaluation Design ..... 5
  - Recruitment and Random Assignment ..... 5
  - Study Design ..... 6
  - Sources of Data and Analysis Strategies ..... 8
- CLI Program Fidelity of Implementation ..... 10
- Effect of the CLI Program on Teacher Literacy Practice ..... 13
- Effect of the CLI Program on Reading Achievement ..... 14
  - Effects on Overall Reading Achievement ..... 14
  - Effects on Individual Reading Skills ..... 17
- Conclusions ..... 18
- References ..... 19
- Notes ..... 21

## Overview

Student literacy has been a major concern in the United States for many years. Research has demonstrated that children from even the most literacy-impooverished backgrounds can learn to read successfully (Snow, Burns, & Griffin, 1998). Yet many of the nation's children leave school without even the most basic reading skills they will need to function in society (Aud et al., 2013). Good teachers, of course, are critical to student achievement (Allen, 2003; Rowan, Correnti, & Miller, 2002; Sanders and Rivers, 1996). To improve teachers' knowledge and skill, federal policymakers have committed significant resources to teachers' professional development. In 2004–05, for example, states and districts spent \$1.5 billion in federal funds on teacher professional development (Birman et al., 2007).

Over the past decade, a large amount of research has been done on teachers' learning and professional development. (See reviews by Borko, 2004; Richardson & Placier, 2001; Supovitz, 2001; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007.) However, only a small percentage of this research focused on the effectiveness of specific methods for providing professional development. Furthermore, research findings have shown mixed results (Biancarosa, Bryk, & Dexter, 2010; Garet et al., 2008; Jacob & Lefgren, 2004; Marsh et al., 2008; Matsumura, Garnier, & Spybrook, 2013). For example, one study (Garet et al., 2008) looked at a program that offered 48 hours of literacy-related professional development to teachers. Half of the teachers received an additional 60 hours of coaching. That study found that the training had no significant effects on teachers' methods or on students' reading achievement for either group of teachers. On the other hand, the Literacy Collaborative coaching model did show significant positive effects on student achievement (Biancarosa, Bryk, & Dexter, 2010). The Literacy Collaborative model included 40 hours of coach-led courses for teachers in the first year, plus 10–12 hours of professional development and coaching in subsequent years.

With funding from a 2010 Investing in Innovation (i3) grant through the U.S. Department of Education, the Children's Literacy Initiative (CLI) implemented an enhanced version of its program that focused on embedded coaching provided by CLI-trained coaches and by trained peers in the school. CLI engaged American Institutes for Research<sup>®</sup> (AIR<sup>®</sup>) to evaluate the impact of this program in kindergarten to grade 2 classrooms. Specifically, AIR evaluated the effect of the CLI program on the classroom environment, teachers' literacy instruction, and students' reading achievement. For the i3 study, the CLI program provided professional development and classroom-embedded coaching in early literacy practices to all teachers within a grade level. The program also identified one teacher in each grade to whom CLI provided additional resources, so that all teachers might benefit from embedded coaching. The study was a randomized controlled trial, conducted in 78 schools across four districts in three states.

## The Children’s Literacy Initiative Program

Founded in 1988 in Philadelphia, CLI is a 501(c)(3) nonprofit organization that provides prekindergarten through third-grade teachers with training and coaching in effective practices for early literacy instruction. CLI’s coaching and professional development program aims to raise student reading achievement by helping teachers establish and maintain literacy-rich classroom environments and develop shared standards of high-quality instruction. The CLI program used in the i3 validation study includes four key features:

- Providing teachers with literacy resources, including book collections such as the independent reading collection and home lending library, to create a literacy-rich classroom environment
- Conducting professional development institutes and seminars to train teachers in strategies and techniques for literacy instruction, followed by classroom-embedded coaching to help teachers apply these strategies in the classroom
- Identifying one model classroom teacher per grade who will receive intensive coaching and support and in whose classroom additional embedded coaching could be provided
- Educating school leaders on how to leverage CLI training to sustain high-quality literacy instruction in the school

The CLI program, which provides continued support over three years, was designed to provide a higher level of literacy-related professional development than elementary teachers typically receive. As shown in Table 1, in the first year of the program, CLI offered participating teachers 98 hours of professional development: 50 hours of coaching, three days of a summer institute, and three days of seminars. Model classroom teachers received an additional 50 hours of coaching. During the program’s second year, CLI offered teachers 49 professional development hours (25 hours of coaching and three days of seminars), with an additional 25 hours of coaching offered to model classroom teachers. In the third year, CLI offered teachers 39 hours of professional development (15 hours of coaching and three days of seminars), with an additional 15 hours of coaching offered to model classroom teachers. Table 1 also lists the professional development hours CLI provided for principals and school-based coaches.

Teachers who remained in the CLI classrooms over the three years could receive up to 186 hours of professional development. Teachers whose classrooms had been designated as model classroom teachers could receive up to 276 hours of professional development, depending on the number of years the teacher served as a model. Teacher mobility was high in the four study districts. To meet the challenge of high mobility, CLI provided supplemental coaching to new teachers and long-term substitute teachers in an effort to catch them up with the other teachers in the school. CLI also identified and trained new model classroom teachers each year, as needed.

**Table 1. Hours of Professional Development (PD) Offered Each Year of the CLI Program**

Program Year	Individual	Summer Institute, Hours	Seminars, Hours	Coaching, Hours	Total PD, Hours	# District-wide Meetings
<b>1</b>	Regular classroom teacher	24	24	50	98	-
	Model classroom teacher	24	24	100	148	6
	Principal	-	-	7	7	6
	School-based coach	16	-	4	20	-
<b>2</b>	Regular classroom teacher	-	24	25	49	-
	Model classroom teacher	-	24	50	74	9
	Principal	-	-	7	7	6
	School-based coach	-	-	4	4	-
<b>3</b>	Regular classroom teacher	-	24	15	39	-
	Model classroom teacher	-	24	30	54	9
	Principal	-	-	7	7	6
	School-based coach	-	-	4	4	-
<b>Total</b>	Regular classroom teacher	24	72	90	186	-
	Model classroom teacher	24	72	180	276	24
	Principal	-	-	21	21	18
	School-based coach	16	-	12	28	-

Source: Study records.

One of CLI’s goals is to work with all early elementary teachers within a school to develop a shared understanding of high-quality early literacy instruction. Coaches build teachers’ deep knowledge and skills by focusing each year on specific instructional areas, such as small-group and individualized instruction. Training institutes and seminars introduce these strategies to teachers. Next, teachers implement these strategies in their classrooms and deepen their understanding through on-site coaching provided by CLI-trained coaches. CLI coaches, who work with the same teachers throughout the year, are trained to use a structured coaching approach: a preconference (coplanning with the teacher), teaching the lesson (observing the teacher, coteaching, or modeling), and a postconference (debrief with the teacher). Teachers learn by doing, observing, and discussing their practice with their coach.

During the first program year, CLI coaches work with teachers to develop a classroom literacy environment that optimizes the use of space, demonstrates the value of reading and writing, and makes use of the high-quality resources CLI has provided. Coaches encourage teachers to establish appropriate procedures, routines, and responsibilities in the classroom, with attention to both teachers' and students' use of language in ways that are accountable to learning.

Next, coaches help teachers implement evidence-based best practices in literacy. The CLI program focuses on the following practices (among others): (1) Message Time Plus<sup>®</sup>—a reading and writing activity for teaching vocabulary, print concepts, phonics, and phonemic awareness; (2) Intentional Read Aloud—a reading and comprehension activity using books chosen to align with lesson objectives and standards; (3) Reading/Writing Workshop—a mini-lesson based on students' needs, followed by students' independent work on self-selected topics and ending with sharing out; and (4) Small Group and Individualized Learning—strategies for differentiating instruction for individuals and small groups.

## i3 Evaluation Design

The i3 evaluation analyzed the effect of the CLI program on the reading achievement of students in kindergarten through second grade. The study team also evaluated the effect of the CLI program on classroom environments and teachers' early literacy practices.

### **Recruitment and Random Assignment**

The study occurred in 78 schools in four school districts located in Illinois, New Jersey, and Pennsylvania. School districts were recruited based on their commitment to teacher development as well as their large populations of low-income and low-performing students. CLI worked with superintendents and district staff to identify schools where the CLI program could have the greatest effect in closing the achievement gap and where CLI had not previously been implemented.

Schools that agreed to participate were randomly assigned to either the CLI program group or the control group.<sup>1</sup> Teachers at the 39 schools in the control group received only the professional development normally provided by their district. Random assignment ensures that the characteristics of schools in the CLI program group and control group are not systematically different. As a result, evaluators can assume that any differences observed between the two groups at the end of the study are caused by the CLI program and not by preexisting differences between the two groups.

At the start of the study, the CLI group and control group were found to be equivalent to one another on baseline reading achievement and on all but one demographic characteristic. The exception was the percentage of students identified as English Language Learners (ELL), with a higher proportion of ELL students in CLI schools than in control schools (estimated difference of 6 percent).<sup>2</sup> This demographic difference was accounted for in all of the CLI impact analyses.

Random assignment was completed in the fall of 2010, one year before CLI implementation began in the study schools and AIR began its data collection.

As an incentive for participation, schools in the control groups received a \$4,000 school library (delivered during the 2010–11 and 2011–12 school years), and schools in the CLI group received coaching and professional development for their third-grade teachers. To minimize any possibility of spillover of CLI training from third-grade teachers to teachers in the study grades (kindergarten through second grade), CLI group school principals agreed in writing that they would not move third-grade teachers into the study grades. An analysis of teacher reports of prior exposure to CLI at the beginning of the study in the fall of 2011 showed no significant differences between the CLI group and the control group, signifying that no substantial spillover occurred.

The study schools served urban populations. All were eligible for federal government funds from Title 1 for serving low-income students. Compared to the national population of Title 1-eligible elementary schools, schools in this study tended to have a larger proportion of students receiving free or reduced-price lunch (84% compared to 59%) and a higher proportion of minority

students, as shown in Table 2. The number of students identified as having a learning disability was comparable to the national average (NCES, 2015).

## Study Design

The i3 evaluation examined the effect of the CLI program on kindergarten through second-grade student achievement during the 2011–12 (Year 1), 2012–13 (Year 2), and 2013–14 (Year 3) school years. The CLI program was phased in over the first two study years to allow time for CLI to scale up its organization and staff to accommodate the large number of CLI teachers in the i3 evaluation. Kindergarten and first-grade teachers were the first to receive services, beginning in Year 1. Regular classroom teachers in these grades who remained in the CLI program for all three study years could receive up to 186 hours of CLI professional development and coaching. Second-grade teachers began receiving services in Year 2. Regular classroom second-grade teachers who remained in the CLI program through Year 3 of the study received up to 147 hours of CLI professional development and coaching.

**Table 2. Characteristics of Study Schools**

School Characteristic	Study Schools	National Population of Title 1 Eligible Elementary Schools
Total number of students	553	449
Race/ethnicity		
White, %	7	50
Hispanic, %	25	26
Black, %	66	16
Asian/Pacific Islander, %	1	3
American Indian/Alaskan Native, %	0	2
Free/reduced-price lunch, %	84	59
English language learners, %	7	n/a
Students with disabilities, %	13	n/a
At or above proficient, third-grade reading 2008–09, %	42	n/a

*Source:* 2010–11 Common Core Data (CCD) and publicly available state-reported data.

Two groups (cohorts) of kindergarten students were followed over the course of the three-year study. The first cohort of students began kindergarten in Year 1, the 2011–12 school year, moving into first grade in Year 2 and second grade in Year 3. The second cohort of students began kindergarten in Year 2, the 2012–13 school year, and were followed into first grade in Year 3. The study team followed two cohorts of students to determine whether the effect of the CLI program was similar across both cohorts. Finding similar results across two different samples of students would increase the team’s confidence in the reliability and validity of the findings.

In addition, by following two sequential groups of students, the study team was able to explore the effects of the CLI program for students whose teachers received different amounts of CLI professional development and coaching. For example, the team compared the Year 1 results for Cohort 1 kindergarten students with the Year 2 results for Cohort 2 kindergarten students who were in the same schools. Table 3 contrasts the number of professional development hours teachers of each of these student cohorts would have had. The team expected to find larger effects of the CLI program on reading achievement in Year 2 for Cohort 2 students, because the CLI teachers would have received professional development and coaching in a broader array of topics and had more opportunity to practice their knowledge with each additional year of the study.

**Table 3. Comparing Differences in Teachers’ Years of Study Participation for Student Outcomes by Student Cohort**

		Kindergarten Impacts	First-Grade Impacts	Second-Grade Impacts
<b>Cohort 1 Students</b> <i>Began kindergarten in the 2011–12 SY</i>	Study year when outcome is measured	Year 1 <i>Spring 2012</i>	Year 2 <i>Spring 2013</i>	Year 3 <i>Spring 2014</i>
	Teachers’ years of participation in study	Up to 1 year	Up to 2 years	Up to 2 years
	Hours of professional development received by CLI teachers	Up to 98 hours	Up to 147 hours	Up to 147 hours
<b>Cohort 2 Students</b> <i>Began kindergarten in the 2012–13 SY</i>	Study year when outcome is measured	Year 2 <i>Spring 2013</i>	Year 3 <i>Spring 2014</i>	—
	Teachers’ years of participation in study	Up to 2 years	Up to 3 years	—
	Hours of professional development received by CLI teachers	Up to 147 hours	Up to 186 hours	—

Student achievement results reflect the CLI program’s effect on all students in regular kindergarten, first-, and second-grade classrooms in the study schools, even though some of those students were not present for the full duration of the study. We will refer to this sample of students as *sample with joiners*.<sup>3</sup> The kindergarten samples with joiners included 4,333 and 4,903 students for Cohorts 1 and 2, respectively. The first-grade samples with joiners included 3,908 and 4,073 Cohort 1 and 2 students, respectively. The second-grade sample with joiners included 3,757 Cohort 1 students. Of the students in the first-grade sample, approximately 70 percent were on the baseline kindergarten rosters, indicating that these students stayed in the study for the full two years (referred to as *stayers*). The remaining 30 percent of students were *joiners* who moved into study schools during the kindergarten or first-grade year. Joiners’ student achievement levels may reflect information learned during their time in study schools and during their time in other locations. Of the second-grade students, 61 percent were on the baseline rosters, and 74 percent were on the first-grade rosters.

Reading achievement results also reflect the fact that some teachers started at the schools after program implementation began and therefore could not participate for the full duration of the study. In Year 2, the team found that 87 percent of kindergarten CLI teachers and 79 percent of first-grade CLI teachers had been in the study both years. By Year 3, only 66 percent of first-grade CLI teachers had participated in all three years of the study, and 76 percent of second-grade CLI teachers had participated in two years of the study. (Control-group teachers had similar rates of stability and mobility as CLI teachers.) CLI teachers who participated in fewer years of the study than their peers received professional development and coaching on fewer topics and had less opportunity to practice what they learned.

It can be useful to report student achievement results for all students in all teacher's classrooms, not just the students and teachers present for the full duration of the study. Such reporting better reflects the reality of student and teacher mobility in schools where CLI typically works, improving the generalizability of this study's results. However, including students and teachers who have not received the full CLI program may result in underestimating the full effect of the CLI program. In addition, analyses that include students who joined the schools after randomization could be biased if students joined CLI schools and control schools for different reasons, leading to nonequivalence in the samples.

The study team found that mobility in CLI and control schools was similar at each point in time. CLI and control-group students who left the study had similar baseline reading achievement levels. Likewise, the team found similar baseline reading achievement between CLI and control-group students who stayed in the study. These analyses suggest that, on the whole, movement into or out of schools did not occur differently for CLI and control students.

To explore whether the inclusion of joiners biased the results, the team analyzed the effect of the CLI program on students who entered study schools at the start of kindergarten and remained in the study schools through the end of kindergarten, first grade, or second grade. We will refer to these students as the *sample of stayers*. The study team included only students with both baseline and outcome data in the analyses. CLI and control-group students were found to be equivalent to one another on reading achievement levels at the start of kindergarten for each of these samples. However, because schools were the unit of random assignment, and all students who joined the schools did so after random assignment, the reading achievement results reported here represent the school-average achievement of students.

School attrition levels were very low during the study. Four schools (two CLI and two control schools) dropped out of the study because the districts closed the schools.<sup>4</sup> Low levels of school attrition, coupled with a lack of differential school attrition, reduce the risk of biased and invalid school-level results.

## **Sources of Data and Analysis Strategies**

During the three-year study, the team selected data sources that would document (a) CLI program implementation, (b) the difference in the amount of professional development services received by CLI teachers and control teachers, (c) the quality of teachers' classroom environment and literacy instruction, and (d) students' reading achievement. Program implementation data

included program participation statistics as well as logs in which coaches recorded the amount of time they spent coaching each teacher and the content focus of the coaching sessions.<sup>5</sup>

The team surveyed teachers at the start of the study and in the spring of every study year to collect data on literacy-related professional development experienced by CLI and control teachers, as well as data on teacher characteristics.

To assess teachers' classroom environment and literacy instruction, the study team observed each teacher's literacy lesson using the Early Language and Literacy Classroom Observation (ELLCO) K–3 Research Tool (Smith, Brady, & Clark-Charelli, 2008).<sup>6</sup> The team randomly selected one CLI teacher and one control teacher from each kindergarten and first-grade classroom to be observed in spring 2013, for a total of 150 teachers. Of those, 130 teachers (including 65 in the CLI program group and 65 in the control group) were observed and included in the impact analysis. The team estimated the effect of the CLI program on teachers' classroom environment and literacy instruction using a model that accounts for the grouping of teachers within schools.<sup>7</sup>

The team assessed students' reading achievement in the fall of the kindergarten year for each cohort of CLI and comparison students. The team also assessed reading achievement in the spring of kindergarten (Cohorts 1 and 2), spring of first grade (Cohorts 1 and 2), and spring of second grade (Cohort 1). In kindergarten classrooms, AIR-trained data collectors administered the Predictive Assessment of Reading (PAR), an individually administered test of prereading skills that takes 12–15 minutes to complete (Wood & Meyer, 2011).<sup>8</sup> The PAR assesses students' letter-word reading, picture vocabulary, phonemic awareness, and rapid naming fluency.

In first- and second-grade classes, AIR-trained data collectors administered the Group Reading Assessment and Diagnostic Evaluation (GRADE), a group-administered multiple-choice assessment that is completed in two sessions lasting a total of 50–70 minutes (Williams, Cassidy, & Samuels, 2001).<sup>9</sup> The GRADE assesses students' word reading/meaning and comprehension.

The team chose both PAR and GRADE on the basis of their demonstrated ability to produce valid, reliable estimates of student's reading achievement (Williams et al., 2001; Wood, Hill, Meyer, & Flowers, 2005; Wood & Meyer, 2011). Researchers estimated the effect of the CLI program on student reading achievement using a two-level hierarchical model that accounts for the grouping of students within schools.<sup>10</sup> These models take into account students' baseline kindergarten achievement scores as well as the percentage of ELL students in the school.<sup>11</sup> The study team assessed more than 80 percent of eligible children each year.

## CLI Program Fidelity of Implementation

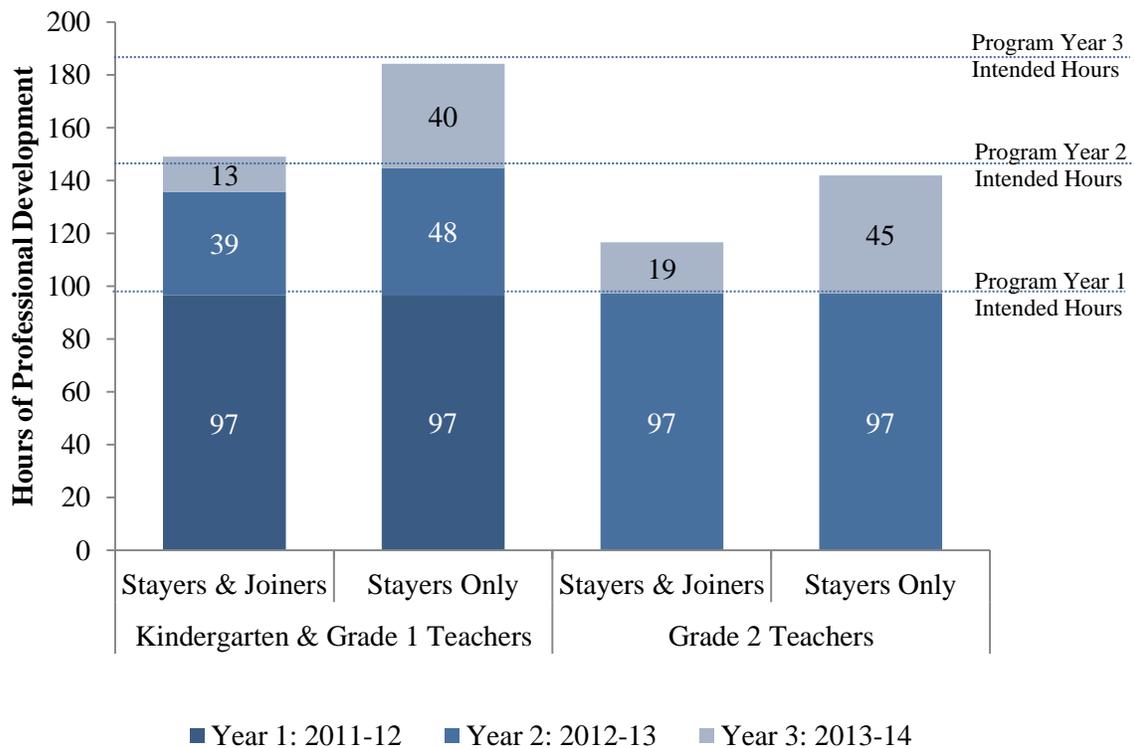
The term *fidelity of implementation* refers to the extent to which implementers adhere to the core principles and practices of a given program. The CLI program was implemented in schools with a high level of fidelity across the three study years.

The study team measured fidelity using an index based on four key components of the CLI program: (a) resources and professional development institutes and seminars provided to teachers; (b) one-on-one coaching provided to teachers; (c) identification of a model classroom, with additional one-on-one coaching and resources provided to the model classroom teacher; and (d) professional development provided to school leaders. The team identified multiple indicators (e.g. the number of coaching hours teachers received) for each program component, and the team gave individuals a score indicating high, moderate, or low fidelity for each indicator.<sup>12</sup> The team combined fidelity data within each school, adding up the fidelity scores for each indicator to obtain a school-level fidelity index for each of the four program implementation components. Overall, components were said to have been implemented with high fidelity if at least 80 percent of schools received a high-fidelity index score. Components were considered to have been implemented with moderate fidelity if at least 50 percent of schools received a high-fidelity index score.

CLI resources for teachers' classrooms and professional development were provided with high fidelity during Years 1 and 2. CLI resources were provided with moderate fidelity during Year 3, when attendance rates at CLI seminars dropped just below the high-fidelity mark. One-on-one coaching for regular classroom teachers, along with coaching and resources provided to model classrooms and their teachers, were provided with high fidelity during all three years of the study. Professional development for school leaders was implemented with moderate fidelity each year. This moderate fidelity was due, in part, to district restrictions on the number of meetings CLI was able to provide for principals, as well as moderate attendance at the meetings offered. Across the key components, the full CLI program was implemented with high fidelity each year of the study.

Teacher turnover during the study did not detract noticeably from CLI's high level of implementation fidelity. Regular classroom teachers who stayed in study schools all possible years participated in nearly all of the intended professional development each year. For example, the second column in Exhibit 1 shows *stayer* kindergarten teachers and first-grade regular classroom teachers. After one year of program participation, teachers received 97 of the intended 98 hours; after the second program year, teachers received 145 of the intended 148 hours; and by the third program year *stayer* teachers received a total of 184 of the intended 186 hours. Including the hours of professional development received by *joiner* teachers (those who moved into study schools after the beginning of the study), as shown in the first column in Exhibit 1, the total average hours received for kindergarten teachers and first-grade regular classroom teachers drops to 146, which is still 80 percent of the total intended hours. The team observed similar patterns for second-grade teachers.

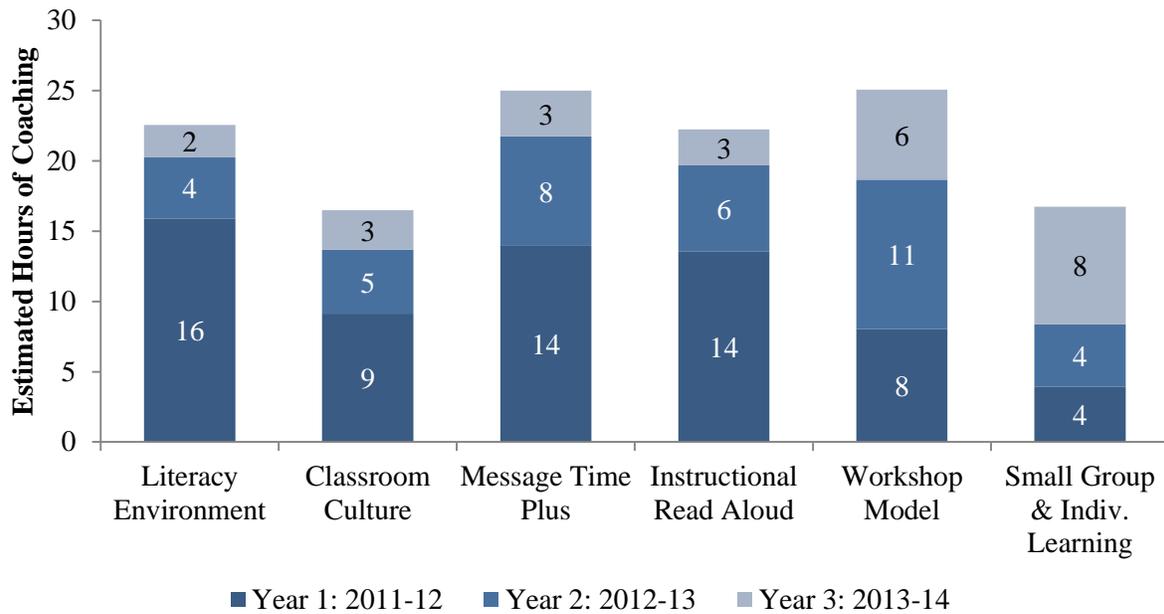
**Exhibit 1. Hours of Professional Development and Coaching Received by CLI Regular Classroom Teachers Each Year**



Source: CLI invoice and attendance records.

The study team also examined whether the coaching sessions were delivered as intended. Coaches completed logs every two weeks. In the logs, they documented the coaching activities in which they engaged with the teacher and the content on which they focused during coaching. The majority of coaches’ logs included planning and debriefing time with the teachers, as well as a classroom-based activity such as observing the teacher, modeling instruction, or coteaching. Exhibit 2 shows the most common content areas of focus during coaching. By the end of Year 3, CLI had delivered at least 15 hours of coaching to kindergarten and first-grade teachers and at least 10 hours of coaching to second-grade teachers in six content areas: literacy environment, classroom culture, Message Time Plus®, instructional read-aloud, workshop model, and small-group and individualized learning. Exhibit 2 shows the distribution of these hours over the three study years for kindergarten and first-grade teachers. On average, coaches and teachers spent at least a few hours on each content area each year, but each year did have content areas that received a greater focus, as intended.

**Exhibit 2. Estimated Coaching Hours Focused on Different Content Areas, by Year, for Kindergarten and First-Grade Teachers**



Source: Coach logs.

As intended, CLI teachers experienced more hours of literacy-related professional development than teachers in control schools, based on teachers' reports of professional development attendance from a study-administered survey. Specifically, CLI teachers reported receiving an estimated 89 more hours of literacy-related professional development and coaching in Year 1, 61 more hours in Year 2, and 59 more hours in Year 3.<sup>13</sup>

Each year of the study, a larger proportion of CLI teachers received professional development with a major focus on the literacy environment, classroom management, and high-quality children's literature compared to professional development that control teachers received. During Year 3, professional development received by CLI teachers tended to have a greater focus on small-group instruction and differentiating instruction than the professional development attended by control teachers; this focus aligns with the content emphasis during that year.

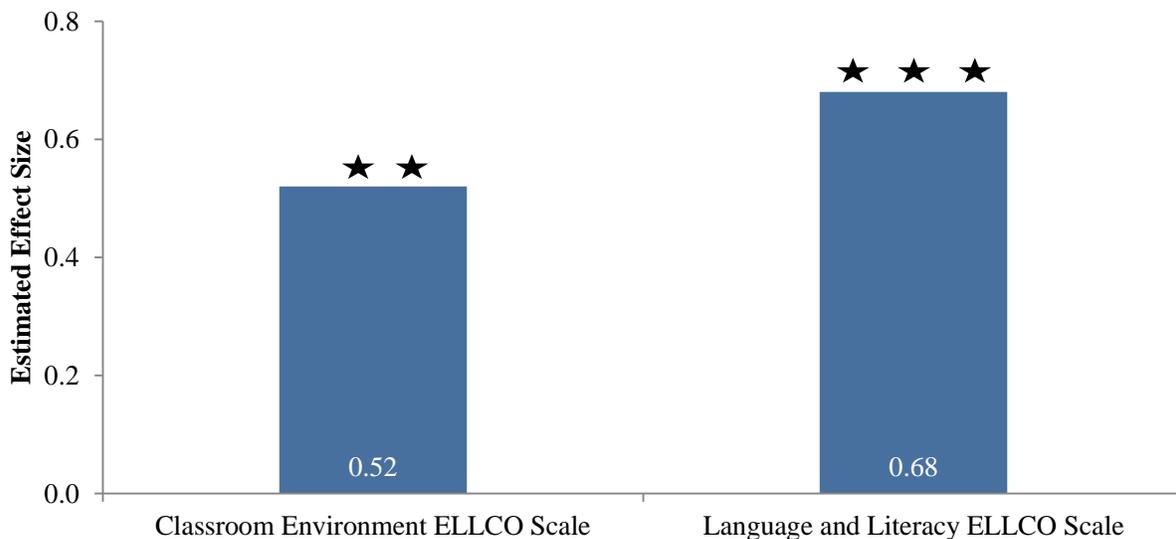
## Effect of the CLI Program on Teacher Literacy Practice

The CLI program had a statistically significant positive effect on kindergarten teachers' and first-grade teachers' classroom environment and language and literacy practices, based on classroom observations conducted by study-trained observers. Exhibit 3 presents these results. These results reflect effects for teachers receiving up to one and a half years of the CLI program.

On a scale of 1 (deficient/minimal evidence) to 5 (exemplary/compelling evidence), teachers in the CLI group had a rating of 3.68 on the classroom environment ELLCO scale, compared to a rating of 3.30 for teachers in the control group. This rating indicates that observers found mostly strong evidence for each indicator of classroom environment. The estimated difference of 0.38 corresponds to an effect size of 0.52 standard deviations ( $p$  value = 0.003), as seen in Exhibit 3.<sup>14</sup> An *effect size* is a common way of calculating program effects; it allows evaluators to compare effects across measures and studies. Effect sizes of about 0.20 or less are considered small, effect sizes around 0.5 are considered medium, and effect sizes of about 0.8 or more are considered large (Cohen, 1988).

On the language and literacy ELLCO scale, teachers in the CLI group had a rating of 3.28, compared to a rating of 2.82 for teachers in the control group. A rating of 3.28 suggests that observers generally found evidence of the indicators of language and literacy practices in CLI teachers' classrooms. The estimated effect of the CLI program on teachers' language and literacy practices represents an effect size of 0.68 ( $p$  value <0.001).

**Exhibit 3. Effect of the CLI Program on Teacher Practice Outcomes**



*Source:* Early Language and Literacy Classroom Observation (ELLCO) Teacher Observations.

*Notes:* Sample size = 65 CLI group and 65 control group teachers. Effect sizes were calculated by using the control group standard deviation.  $P$ -values are based on two-tailed  $t$  tests: ★★  $p$  < 0.01; ★★★  $p$  < 0.001.

## Effect of the CLI Program on Reading Achievement

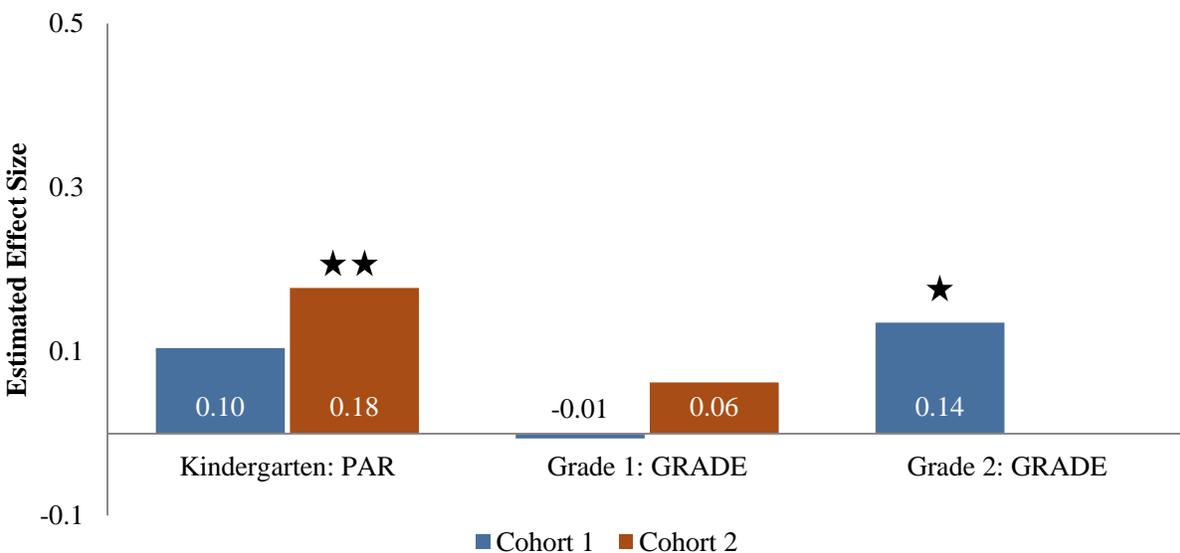
The CLI program had a positive and significant effect on average reading achievement in kindergarten and second grade, but no significant effect on average first-grade reading achievement.

This section describes the effect of CLI on reading achievement by grade for two separate cohorts of students. Cohort 1 students began kindergarten in the 2011–12 school year, and Cohort 2 students began kindergarten in the 2012–13 school year. Kindergarten Cohort 1 outcomes were measured while schools were participating in their first year of the study, and teachers in CLI schools had participated in up to one year of the CLI program. Kindergarten Cohort 2 outcomes were measured when schools were participating in their second year of the study, and teachers in CLI schools had participated in up to two years of the CLI program. First-grade Cohort 1 outcomes were measured while schools were participating in their second year, and Cohort 2 first-grade outcomes were measured while schools were participating in their third year. Second-grade Cohort 1 outcomes were measured while schools were participating in their third and final year of the study.

### Effects on Overall Reading Achievement

Exhibit 4 shows the impact of the CLI program on average grade-level reading achievement by cohort. In kindergarten classrooms, the study team found effect sizes of 0.10 in Year 1 for Cohort 1 ( $p$  value = 0.066) and 0.18 in Year 2 for Cohort 2 ( $p$  value = 0.002).<sup>15</sup> An effect size of 0.18 corresponds to an improvement index of 7 percent, meaning the percentile rank of the average control student would increase from the 50th percentile to the 57th percentile if the student had been in the CLI group.<sup>16</sup>

**Exhibit 4. Effect of the CLI Program on Overall Reading Achievement**



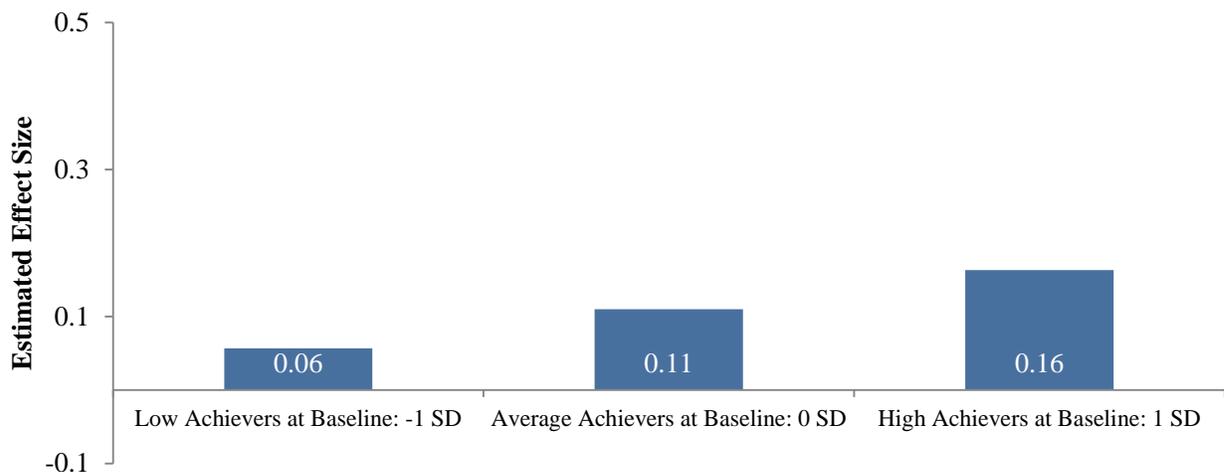
*Source:* Study-administered kindergarten reading assessment Predictive Assessment of Reading (PAR) and first- and second- grade reading assessment Group Reading Assessment and Diagnostic Evaluation (GRADE).

*Notes:* Effect sizes were calculated by using the control group standard deviation.  $P$ -values are based on two-tailed  $t$ -tests: \*  $p < 0.05$ ; \*\*  $p < 0.01$ .

Comparing the size of the effects across the two cohorts, the study team found that Year 2 Cohort 2 effects were significantly larger than Year 1 Cohort 1 effects ( $p$  value  $<0.001$ ). This difference is likely due to the difference in years of study participation. At the time of Year 1 Cohort 1 results, CLI teachers had received an average of 97 hours of professional development and coaching. The larger results found for Cohort 2 in Year 2 were measured after CLI regular classroom teachers had received an additional 39 professional development and coaching hours, on average, and they had had an additional year to practice enacting their acquired knowledge. In addition, by Year 2, model classroom teachers had been identified and developed for two years; thus, they may have been a more valuable resource to their peers than they were in Year 1.

Although the CLI program did not have a statistically significant effect for all students in Year 1, the program had a significant differential effect for students with different baseline achievement levels.<sup>17</sup> Findings suggest that, in Year 1, the CLI program was more effective for students with higher baseline achievement, compared to students with average or lower baseline achievement. Exhibit 5 shows estimated effect sizes for three types of students representing high, average, and low baseline achievers. CLI had an effect size impact of 0.16 for higher-achieving students, whereas CLI only had an effect of 0.06 for lower-achieving students. By Year 2, however, CLI had similarly sized effects for all students, regardless of their incoming achievement level.

**Exhibit 5. Differential Effects of the CLI Program on Kindergarten Reading Achievement in Year 1 for Cohort 1 Students With Different Levels of Baseline Achievement**



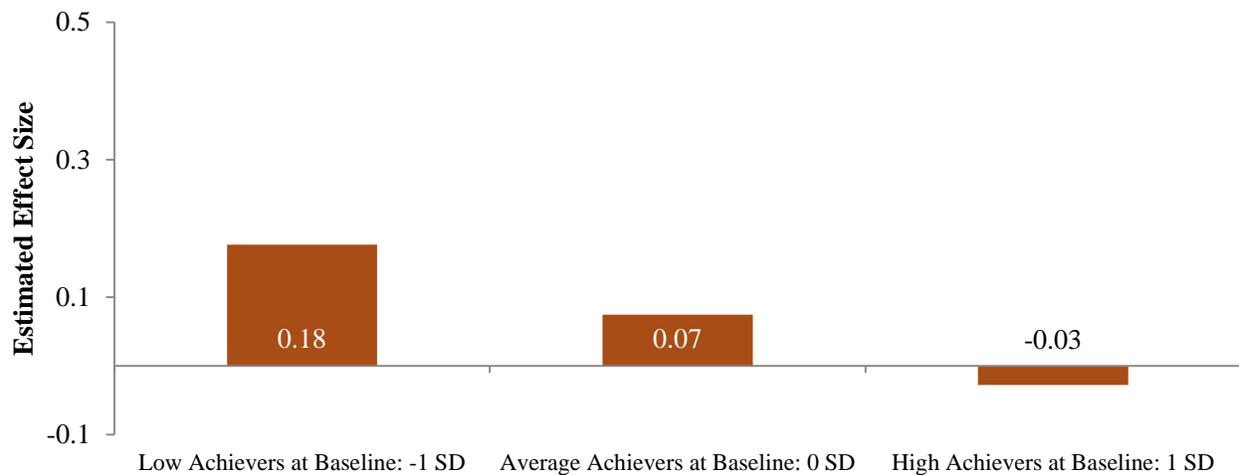
Source: Study-administered kindergarten reading assessment, Predictive Assessment of Reading (PAR).

Notes: Kindergarten cohort 1 sample size = 39 CLI schools and 38 control schools; 2,275 CLI students and 2,058 control students. Effect sizes were calculated using the control group standard deviation of 12.32.

In first-grade classrooms, the study team found that the CLI program had no significant effect on average reading achievement in either cohort. However, as in kindergarten, the team found a trend in the size of effects. The team observed larger, though nonsignificant, effects in Year 3 for Cohort 2 (effect size of 0.06,  $p$  value = 0.415) compared to a nearly zero effect estimate in Year 2 for Cohort 1 (effect size of -0.01,  $p$  value = 0.931). Additional analysis found a significant differential effect of CLI on first-grade achievement in Year 3.<sup>18</sup> These analyses suggest that the CLI program did affect first-grade reading achievement for students who were low-achieving at

baseline. As Exhibit 6 shows, low-achieving students who were below average at baseline had an estimated effect size of 0.18.

### Exhibit 6. Differential Effects of the CLI Program on First-Grade Reading Achievement in Year 3 for Cohort 2 Students With Different Levels of Baseline Achievement



Source: Study-administered first-grade reading assessment, Group Reading Assessment and Diagnostic Evaluation (GRADE).

Notes: First-grade cohort 2 sample size = 38 CLI schools and 37 control schools; 2,189 CLI students and 1,884 control students. Effect sizes were calculated by using the control group standard deviation of 17.51.

The effect size of the CLI program on second-grade average reading achievement was 0.14 ( $p$  value = 0.038). Schools were in their third year of study participation when second-grade student outcomes were measured, but CLI teachers had received only up to two years of the CLI program. This effect represents a difference of approximately two months of reading achievement growth, or a 3 percentile improvement (on average, the CLI group scored at the 38th percentile compared to the control group, which scored at the 35th percentile).<sup>19</sup> Exploratory analysis shows that CLI had a similar effect for all students, with no differential effect based on incoming kindergarten baseline knowledge.

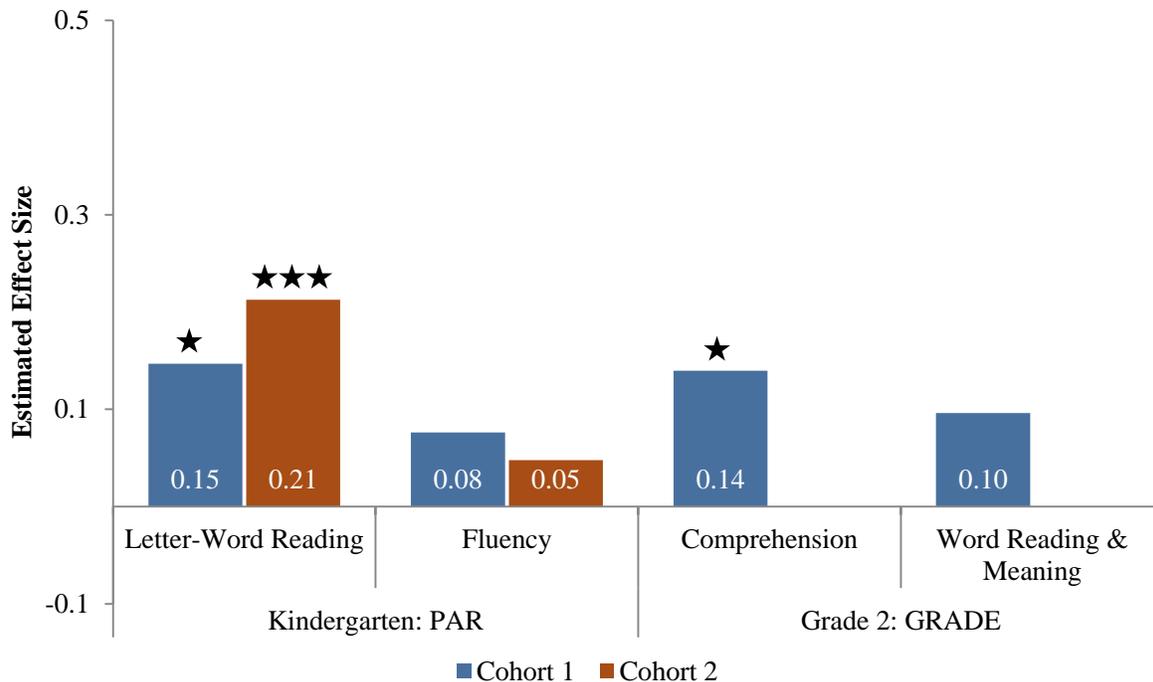
The student achievement results presented above represent schools with both *stayers* (students who stayed in the study since the beginning of kindergarten) and *joiners* (students who joined the study after the start of kindergarten)—a sample that reflects the reality of student mobility in schools where CLI typically works. Inclusion of joiners, however, could bias or underestimate the program's effect. Results from analyses restricting the student sample to only *stayers* with baseline data were very similar to results including *joiners*, with positive and significant effects on average kindergarten reading achievement in both cohorts (effect sizes of 0.12 and 0.17, respectively), no significant effect on first-grade reading achievement, and a significant positive effect on second-grade average reading achievement (effect size 0.13). The similarity in size and significance of results indicates these results are robust, and the results for the more representative sample including *joiners* are valid.

## Effects on Individual Reading Skills

In addition to examining the effect of the CLI program on overall reading achievement, the study team explored the effect of CLI on individual reading skills. We assumed the CLI program would have a greater effect on more advanced skills that build on prereading skills. Four skills were measured in kindergarten with the PAR, and the team hypothesized that CLI would have an effect on two of these skills: letter–word reading and fluency. Two skills were measured in first and second grade with the GRADE; the team hypothesized that both comprehension and word reading and meaning would be affected in first grade and that comprehension would be affected in second grade.

As Exhibit 7 shows, the CLI program had a positive and significant effect on kindergarten average letter–word reading, with an effect size of 0.15 in Year 1 for Cohort 1 ( $p$  value = 0.010) and 0.21 in Year 2 for Cohort 2 ( $p$  value < 0.001).<sup>20</sup> As seen with the overall reading achievement results, Year 2 effects were larger than Year 1 effects, indicating that the extra year of services provided in CLI schools led to greater gains in letter–word reading skills. The team found CLI had no significant effect on kindergarten average fluency or on first-grade reading skills. Results for second grade show CLI had a significant positive effect on comprehension skills, with an effect size of 0.14 ( $p$  value = 0.037) representing a two-month gain in comprehension skills.

**Exhibit 7. Effect of the CLI Program on Individual Reading Skills**



*Source:* Study-administered kindergarten reading assessment, Predictive Assessment of Reading (PAR), and second-grade reading assessment, Group Reading Assessment and Diagnostic Evaluation (GRADE).

*Notes:* Effect sizes were calculated by using the control group standard deviation.  $P$ -values are based on two-tailed  $t$  tests: ★  $p$  < 0.05; ★★★  $p$  < 0.001.

## Conclusions

The findings from this study demonstrate that the CLI program produces substantial effects on teachers' classroom environment and literacy practices which, in turn, lead to measurable effects on average reading achievement in early elementary grades. These results are based on a large, well-designed, well-implemented study that took place across four districts and a large number of schools. This research adds to the evidence that providing teachers with coaching and professional development focused on pedagogical content knowledge of literacy can lead to changes in teachers' practice and students' reading achievement.

One reason why this study had positive findings, whereas studies of similar programs have not, may be that the CLI program is aligned with the core features of effective professional development identified by Garet and colleagues (2001): (1) a focus on content knowledge, (2) opportunities for active learning, and (3) coherence with other learning activities. The majority of CLI's professional development and coaching hours focus on literacy-related pedagogical content knowledge, and the in-classroom coaching sessions provide opportunities for active learning and the enactment of the knowledge into practice. In addition, teachers in CLI schools reported that CLI was strongly aligned with what teachers learned in non-CLI professional development, their own goals, state and district standards, and their schools' improvement plan.

Study results indicate that the size of the effect on average reading achievement increases with additional years of CLI program participation. This increased effect over time could be attributed to a number of factors, including the additional content areas covered, the longer time teachers had to work on aligning their practices with what they learned during professional development and coaching, or the availability of model classrooms and teachers who increased opportunities for in-person coaching for the other teachers in their grade level. This pattern of increased effects with increased CLI program participation could be particularly important in first grade, where the study found no overall effect, but did find results suggesting that CLI can positively affect first-grade students who have low reading achievement when they enter school.

In this study, the amount of literacy-related services provided by CLI far surpassed the amount of services typically provided to teachers in the study districts. This study provides evidence that an intensive professional development and coaching program can be implemented with fidelity over multiple years and produce positive effects on teacher practice and reading achievement in early elementary grades, despite common challenges such as teacher and administrator turnover; changes in district leadership, curricula, and standards; and the logistical complexity of providing teachers with 50–100 hours of services yearly.

## References

- Allen, M. (2003). *Eight questions on teacher preparation: What does the research say?* Denver, CO: Education Commission of the States (ECS). Retrieved from <http://www.ecs.org/html/educationissues/teachingquality/tpreport/home/index.asp>
- Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., & Zhang, J. (2013). *The condition of education 2013* (NCES 2013-037). Washington, DC: U.S. Department of Education, National Center for Education Statistics. Retrieved from <http://nces.ed.gov/pubsearch>
- Biancarosa, G., Bryk, A. S., & Dexter, E. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal*, *111*(1), 7–34.
- Birman, B., Le Floch, K. C., Klekotka, A., Ludwig, M., Taylor, J., Walters, K., . . . Yoon, K.-S. (2007). *State and local implementation of the No Child Left Behind Act: Vol. 2. Teacher quality under NCLB: Interim report*. Washington, DC: U.S. Department of Education; Office of Planning, Evaluation and Policy Development; Policy and Program Studies Service. Retrieved from <http://eric.ed.gov/?id=ED497970>
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Earlbaum.
- Garet, M., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., . . . Szejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from <http://ies.ed.gov/ncee/pdf/20084030.pdf>
- Garet M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915–945.
- Jacob, B. A., & Lefgren, L. (2004). The impact of teacher training on student achievement: Quasi-experimental evidence from school reform efforts in Chicago. *Journal of Human Resources*, *39*(1), 50–79.
- Marsh, J. A., McCombs, J. S., Lockwood, J. R., Mantorell, F., Gershwin, D., Naftel, S., . . . Grego, A. (2008). *Supporting literacy across the sunshine state: A study of Florida middle school reading coaches*. Santa Monica, CA: RAND Education. Retrieved from [http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND\\_MG762.pdf](http://www.rand.org/content/dam/rand/pubs/monographs/2008/RAND_MG762.pdf).

- Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2013). Literacy coaching, reading comprehension instruction and student achievement: A mediation model. *Learning and Instruction, 25*, 35–48.
- National Center for Education Statistics, Institute of Education Sciences. (2015). Fast facts. Students with disabilities (NCES 2015-011). Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences. Retrieved from <http://nces.ed.gov/fastfacts/display.asp?id=64>.
- Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed.) (pp. 905–947). New York, NY: Macmillan.
- Rowan, B., Correnti, R., & Miller, R. J. (2002). What large-scale survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. *Teachers College Record, 104*(8), 1525–1567.
- Sanders, W. L., & Rivers, J. C. (1996). Cumulative and residual effects of teachers on future student's academic achievement. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.
- Smith, M. W., Brady, J. P., & Clark-Chiarelli, N. (2008). *Early language and literacy classroom observation K–3 tool, Research edition*. Baltimore, MD: Paul H. Brookes.
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academies Press.
- Supovitz, J. A. (2001). Translating teaching practice into improved student performance. In S. H. Fuhrman (Ed.), *From the Capitol to the classroom: Standards-based reform in the states* (pp. 81–98). (100th Yearbook of the National Society for the Study of Education, Part II). Chicago, IL: University of Chicago Press.
- Williams, K. T., Cassidy, J., & Samuels, S. J. (2001). *GRADE teacher's scoring and interpretive manual level 1*. San Antonio, TX: Pearson Education.
- Wood, F. B., Hill, D. F., Meyer, M. S., & Flowers, D. L. (2005). Predictive assessment of reading. *Annals of Dyslexia, 55*(2), 193–216.
- Wood, F., & Meyer, M. (2011). *PAR predictive assessment of reading: Teacher's manual and starter kit*. Winston-Salem, NC: Child's Mind Publishing.
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>

## Notes

---

<sup>1</sup> In three of the districts, schools were grouped into two or three blocks of schools with similar characteristics (e.g., geographic location, student demographics, past academic performance), and one-half of the schools within each block were randomly assigned to the CLI program group. The use of blocks helped to ensure balance between schools in the CLI program group and control group on characteristics that might be related to student achievement.

<sup>2</sup> This significant difference is driven by differences in ELL levels at schools in two of the study districts. At the time of randomization, an odd number of schools in each district were identified as having high proportions of ELL students. Randomization in both districts put the “odd number” school with high proportions of ELL students into the CLI group, resulting in nonequivalence at baseline.

<sup>3</sup> Student samples were intent-to-treat, meaning that students were grouped with the first school they attended upon entry into the study.

<sup>4</sup> Due to the intent-to-treat nature of our analyses, students who previously attended these closed schools who moved into other study schools were still included in all impact analyses as members of the closed school.

<sup>5</sup> Coach logs were completed electronically as part of the CLI time-reporting and invoicing system. Invoices could not be submitted until the log was complete, leading to nearly 100 percent log-completion rates.

<sup>6</sup> The ELLCO tool consists of 18 items, each rated on a scale of 1 (deficient/minimal evidence) to 5 (exemplary/compelling evidence). The 18 items are organized into two subscales: the general classroom environment subscale (7 items) and the language and literacy subscale (11 items). Additional information about the ELLCO can be found here: <http://products.brookespublishing.com/Users-Guide-to-the-Early-Language-and-Literacy-Classroom-Observation-Tool-K-3-ELLCO-K-3-Research-Edition-P398.aspx>.

<sup>7</sup> The Impact model included randomization blocks as fixed effects and school-average percentage of students identified as ELL. Impact estimates were calculated for each district within a single model, and the overall difference between the CLI and control group was computed as a precision-weighted average across the four districts.

<sup>8</sup> Students respond to questions orally, and data collectors record students’ responses. Two comparable forms of the PAR were administered, one in the fall and one in the spring of the kindergarten year for each cohort of students. Additional information about the PAR can be found here: <http://childsmindpublishing.org/par/content/about-par>.

<sup>9</sup> Data collectors prompt students to complete questions, using information in their test booklets as well as information provided verbally by the data collector. Each question has four answer choices, and students indicate their answers in their test booklet. Additional information about the GRADE can be found here: <http://www.pearsonassessments.com/learningassessments/products/100000646/group-reading-assessment-and-diagnostic-evaluation-grade-grade.html>.

<sup>10</sup> Impact estimates were obtained for each district and then averaged across the four districts, weighting each district’s estimate in proportion to the number of control schools in the study sample. This approach used the data for all four districts in a single analysis.

<sup>11</sup> Randomization block fixed effects were also included in the model. In cases with missing covariate measures, the missing data were replaced with district means, and a dichotomous variable indicating the missing status of a given covariate for each observation was added to the impact analysis model.

<sup>12</sup> One example of a fidelity indicator is the number of coaching hours teachers received. Teachers received a high-fidelity score if they received in their first program year 46–50 coaching hours, a moderate-fidelity score for 31–45 coaching hours, and a low-fidelity score for 30 coaching hours or fewer.

---

<sup>13</sup> Estimated differences are reported for model classroom and regular classroom teachers combined. As expected, model classroom teachers did report more hours for coaching than regular classroom teachers reported.

<sup>14</sup> Control group standard deviations were 0.73 for the Classroom Environment ELLCO Score and 0.67 for the Language and Literacy ELLCO Score.

<sup>15</sup> Control group standard deviations used to calculate effect sizes were 12.32 and 11.98 for kindergarten cohort 1 and 2 PAR total test score, respectively; 17.08 and 17.51 for first-grade cohort 1 and 2 GRADE total test score, respectively; and 15.46 for second-grade cohort 1 GRADE total test score. Kindergarten cohort 1 sample size = 39 CLI schools and 38 control schools; 2,275 CLI students and 2,058 control students. Kindergarten cohort 2 sample size = 38 CLI schools and 37 control schools; 2,305 CLI students and 2,012 control students. Grade 1 cohort 1 sample size = 39 CLI schools and 38 control schools; 2,095 CLI students and 1,813 control students. Grade 1 cohort 2 sample size = 38 CLI schools and 37 control schools; 2,189 CLI students and 1,884 control students. Grade 2 cohort 1 sample size = 39 CLI schools and 38 control schools; 1,945 CLI students and 1,812 control students.

<sup>16</sup> The improvement index was calculated by following the methods used by What Works Clearinghouse (see <http://ies.ed.gov/ncee/wwc>). The change in percentile rank is based on the outcome distribution in the control group.

<sup>17</sup> Results are based on a model similar to the main impact model that included an interaction term between baseline achievement and the treatment indicator. The treatment by baseline achievement interaction estimate was 0.65 with a *p* value of 0.020.

<sup>18</sup> The treatment by baseline–achievement interaction estimate was -1.53 with a *p* value of 0.012.

<sup>19</sup> Grade equivalence and percentile improvements are based on the GRADE normative sample, which is representative of the racial and economic distribution of students in the United States. This study’s sample includes a higher proportion than in the norming sample of minority students and students who qualify for free and reduced-price lunch.

<sup>20</sup> Control-group standard deviations used to calculate effect sizes were 16.50 and 13.73 for kindergarten cohort 1 PAR letter–word reading and fluency subscores, respectively; 15.75 and 13.40 for kindergarten cohort 2 PAR letter–word reading and fluency subscores, respectively; and 13.98 and 16.58 for second-grade GRADE comprehension and vocabulary subscores, respectively. Kindergarten cohort 2 sample size = 38 CLI schools and 37 control schools; 2,305 CLI students and 2,012 control students. Grade 2 cohort 1 sample size = 39 CLI schools and 38 control schools; 1,945 CLI students and 1,812 control students.