Longitudinal Impacts of the Children's Literacy Initiative Professional Development, Coaching, and Model Classroom Intervention on Early Literacy Achievement

Julia Parkinson, American Institutes for Research
John Meakin, American Institutes for Research
Terry Salinger, American Institutes for Research

Abstract Body

Background / Context

Student achievement in literacy has been a focal concern in the United States for many years. Rigorous research has demonstrated that children from even the most literacy-impoverished backgrounds can learn to read successfully (Snow, Burns and Griffin 1998). Yet many of the nation's children leave school without even the rudimentary 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). Improving teachers' knowledge and skill that leads to improved student achievement, particularly in the early grades, can place children on an improved trajectory that can have long-term impacts on life outcomes.

Over the past decade, a large body of literature has emerged on teacher learning and PD (for reviews, see Borko 2004; Richardson and Placier 2001; Supovitz 2001; Yoon et al. 2007). Although the literature is vast, relatively little of it is based on rigorous research about the effectiveness of different methods for providing PD. While there have been a few high-quality randomized trials (Garet et al, 2008; Matsumura et al, 2013) and some quasi-experiments(Jacob & Lefgren, 2004; Biancarosa, Bryk, & Dexter, 2010), in general, more work has been done on the associations between PD and changes in teaching practice than on the association with improved student achievement. Of the more rigorous recent research, the impact of PD and coaching programs on student achievement is mixed.

Purpose / Objective / Research Question / Focus of Study

With funding from an Investing in Innovation (i3) validation grant through the U.S. Department of Education, the Children's Literacy Initiative engaged the American Institutes for Research (AIR) to conduct an independent impact evaluation. The i3 impact evaluation was designed to evaluate the effect of the CLI Model Classroom (MC) coaching and professional development program on teacher's classroom environment and literacy instruction, as well as the effect of the CLI program on student literacy achievement. The CLI program provides three years of professional development and coaching in literacy practices to all teachers in kindergarten through second grade and provides additional resources and coaching to one teacher per grade, in an effort to develop that teacher into a model teacher for the rest of the teachers in the grade. By providing coaching and PD services for three years to teachers, CLI hopes to create a sustained and lasting impact on teachers' literacy practices and classroom environment, and in turn, impact the current and future kindergarten through second grade students in these teachers' classrooms. This study addressed the following research questions: (1) What impact did the CLI Model Classroom treatment have on teacher classroom environment and literacy practices in early literacy? And (2) What impact did the CLI Model Classroom treatment have on student achievement in early literacy? Results from the second year of the study will be reported.

Setting

This study was conducted in four large school districts located in the Midwest and Eastern United States. All of the school districts have large populations of low-income and low-performing students, and all outlined the need for increased literacy-based professional development in their strategic plans. The schools in the study districts were largely urban (3 of the 4 were located in large or mid-sized cities) and served many low-income students (all of the schools were Title I schools, and on average 83% of their students were eligible for free or

reduced-price lunch). The number of students identified as having a learning disability was comparable to the national average.

Population / Participants / Subjects

Seventy-eight schools were recruited within the four school districts. By the second year of the study, 3 schools had closed, leaving 38 treatment and 37 control schools. In the second year of the study, all regular-education kindergarten and first grade teachers were members of the teacher sample. Kindergarten and first grade teachers were participating in their second year of the study, and for treatment teachers, receiving their second year of the CLI program. The second-year teacher sample consists of the 218 kindergarten and 228 first grade teachers in the study schools who were teaching eligible classes in the spring of the second year of the study (spring 2013). All teachers in the second-year impact sample had at least a bachelor's degree and 45 percent had a master's degree. Over four-fifths had had four or more years of teaching experience when they entered the study.

All kindergarten and first grade students in the teachers' regular classes became members of the student sample. First grade students, referred to as cohort 1 students, were included in the analysis sample if they had outcome data in the spring of 2013. These students included both students who had been in the study kindergarten students in the study in the 2011-12 school year and were followed through first grade (their second year of the study) as well as students who had entered the study during the 2012-13 school year. Cohort 1 first grade students in the treatment group could be in classrooms with teachers who were participating in either their first or second year of the CLI program. Kindergarten students in the second year of the study (2012-13 school year), referred to as cohort 2, were in classrooms with kindergarten teachers participating in either their first or second year of the study. The second-year student impact analysis sample consists of 4321 kindergarten and 3937 first grade students with end of year data. On average, 7 percent of students were Limited English Proficient, 84% of students were eligible for free or reduced lunch, 66% of students were Black/African American, and 25% of students were Hispanic.

Intervention / Program / Practice

The CLI program delivered across the two years of the study was designed to establish literacy-rich environments, shared standards of practice, and develop model teachers who could mentor and support their colleague teachers. In the first year, the PD included three summer institute days, three seminar days scheduled during the school year, and 50 hours of classroom-embedded coaching for all eligible kindergarten and first grade teachers in treatment schools. Teachers also receive literacy resources, including book collections (e.g. independent reading collection, read aloud collection, home lending library) and Message Time Plus® materials. After the first few months of coaching, CLI coaches worked with the principal to select one teacher in each grade to serve as a Model Classroom (MC) teacher. The MC teachers are selected on the basis of their willingness to adopt CLI practices, be a mentor, and open their classroom to visits from their peers, administration, or outside funders. MC teachers received additional classroom-embedded coaching and support, a stipend, and additional resources for their classrooms. The CLI program in year 1 also included principal meetings, principal coaching, and coaching meetings for school-based coaches. (See Table 1.)

In the second year, the PD included three seminar days and 25 hours of classroom-embedded coaching for teachers and continued additional support for MC teachers, principals and school-based coaches. The overall dosage was lower in the second year compared to the first year.(See

Table 2.) The intended dosage of content-focused PD in each year (98 hours in the first year and an additional 49 hours in the second year for colleague teachers) was higher than the dosage of literacy-related PD that most elementary teachers typically receive in a single year.

Research Design

The CLI i3 Impact Study used a cluster random assignment design in which 78 schools within each of four districts were randomly assigned either to the treatment or control condition. In three of the districts, schools were grouped into two or three blocks of schools with similar characteristics (e.g., geographic location, demographic characteristics, past academic performance), and half the schools within each block were randomly assigned to the treatment group. The original school sample consisted of 39 treatment and 39 control schools. All teachers in the 78 schools continued to receive the professional development normally provided by the district, with treatment teachers also receiving the CLI professional development, coaching and Model Classroom program. All kindergarten and first-grade teachers and students in the study schools were invited to participate in data collection activities during years 1 and 2.

Data Collection and Analysis

Teachers' classroom environment and literacy instructional practices were measured by the Early Language and Literacy Classroom Observation (ELLCO) Tool. AIR-trained observers visited the classrooms of one randomly-selected teacher per grade in each school once in the spring of 2013. Eighteen items of the ELLCO tool were combined into two subscales: general classroom environment and language and literacy.

Measures of student literacy achievement were gathered for kindergarten and first grade students over the first two years of the study. All kindergarten through second grade teachers and students were invited to participate in the study. In year 1 of the study, all regular education kindergarten students (cohort 1) were administered the Predictive Assessment of Reading (PAR), an individually-administered test of pre-reading skills that assessed students' word reading, vocabulary, phonemic awareness, and rapid naming fluency, was used with kindergarten students (Wood, 2005, Wood 2011). The PAR was administered in the fall and spring by AIR-trained test administrators. In year 2 of the study, a second cohort of kindergarten students were administered the PAR in the fall and spring and all first grade students (including students who were in the school in the previous year of the study and students who had moved into the students) were tested in the fall and spring using the Group Reading Assessment and Diagnostic Evaluation (GRADE), a group-administered test of reading skills that assessed students' word reading and meaning and comprehension(Williams, Cassidy, & Samuels, 2001). Both student assessments demonstrate strong validity and reliability.

The basic strategy for the impact analysis was to estimate the difference in outcomes between the treatment and control groups, adjusting for the blocking used in random assignment and for teacher- and student-level covariates. Because random assignment was conducted separately within each of the four school districts participating in the second year of the study, the study comprised of four random assignment experiments. To obtain the impact estimates, we pooled the data for all four two-year districts in a single analysis, treating the districts as fixed effects. Separate program 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 treatment schools from the district in the study sample. Findings in this report therefore represent the impact on the performance of teachers and students in the average treatment school in the districts.

For outcomes that were measured at the teacher level, a two-level hierarchical linear model was used, with teachers nested within schools. Every teacher within a district was weighted equally (i.e., an implicit weight of 1 was applied for each teacher). For the student achievement outcomes, we used a two-level hierarchical model, in which students were nested within schools. Each student in the sample was weighted equally. The covariates in the student achievement model included student's baseline kindergarten PAR score and a single school-level covariate—the school average percent of students identified as English language learners (ELL).

Findings / Results

During the first two years of the CLI i3 impact evaluation, the AIR study team found that CLI was, overall, implemented with fidelity. CLI had high levels of implementation for their three main program components: 1) resources and professional development seminars and trainings provided to all CLI teachers, 2) coaching provided to all CLI teachers, and 3) additional coaching and professional development seminars and trainings provided to Model Classroom teachers. CLI had moderate fidelity to the planned professional development provided to school-based coaches, and low fidelity of implementation for principal coaching and professional development. As designed, a large service contrast was observed in the amount and type of professional development and coaching received by treatment and comparison teachers, with treatment teachers reporting that they had received substantially more literacy-focused professional development. (See Table 3.)

The study found significant positive results on the impact of the CLI program on teacher's language and literacy practices and on classroom environment, as measured by the ELLCO observation tool The estimated standardized effect size of the CLI Model Classroom program on language and literacy practices was 0.68 standard deviations ($p<0.001$), and the impact on classroom environment was 0.52 standard deviations ($p=0.003$), as measured by the ELLCO. Effect sizes show the strength of differences between two groups, in this case the CLI and non-CLI teachers. (See Table 4.)

Separate impact estimates were calculated for each subscale of the kindergarten and first grade assessments. Cohort 2 kindergarten students in CLI classrooms significantly outperformed control students on one of four PAR subscales—letter-word reading, with an estimated standardized effect size of 0.17 standard deviations ($p=0.007$) in the second year. This replicated the significant results found for letter-word reading with cohort 1 kindergarten students in the 2011-12 school year. The impact of CLI on cohort 2 kindergarten students' PAR Early Reading Skills total score was positive and significant, with an estimated standardized effect size of 0.18 ($p=0.003$). No significant impact was found for CLI on cohort 1 first grade students' literacy achievement on either GRADE subscale. (See Tables 5, 6, and 7).

Conclusions

The impact of CLI on kindergarten word reading for two cohorts of students is promising evidence of the impact of prolonged participation in the CLI program. The significant impact of CLI on student achievement after only one year stands in contrast to a number of studies of professional development programs that have found no significant effects. However, the lack of effect in first grade despite high levels of implementation could indicate poor alignment of the coaching content with student learning objectives in that grade. The final study year will provide additional results for first and second grade students to determine if three years of teacher participation in the CLI program leads to positive impacts on achievement grades.

Appendices

Appendix A. References

Allen, M. (2003). Eight Questions on Teacher Preparation: What Does the Research Say. Education Commission of the States (ECS). Retrieved Jan 2014 from: http://www.ecs.org/html/educationissues/teachingquality/tpreport/home/index.asp.

Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., and Zhang, J. (2013). *The Condition of Education 2013* (NCES 2013-037). U.S. Department of Education, National Center for Education Statistics. Washington, DC. Retrieved Jan. 2014 from http://nces.ed.gov/pubsearch.

Biancarosa, G., Bryk, A. S., and Dexter, E. (2010). Assessing the Value-Added Effects of Literacy Collaborative Professional Development on Student Learning. *Elementary School Journal*, 111 (1), 7-34.

Borko, H. (2004). "Professional Development and Teacher Learning: Mapping the Terrain." *Educational Researcher*, 33(8): 3–15.

Garet, M., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (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.

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*, University of Wisconsin Press, vol. 39(1).

Matsumura, L.C., Garnier, H.E., & Spybrook, J. (2013). Literacy coaching, reading comprehension instruction and student achievement: A mediation model. *Learning and Instruction, 25.*

Richardson, V., and Placier, P. (2001) Teacher Change. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th Ed.). New York: Macmillan, 905–947.

Rowan, B., R. Correnti, and R.J. Miller. What large-scale, survey research tells us about teacher effects on student achievement: Insights from the Prospects study of elementary schools. Teachers College Record , 2002, 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.

Snow, C.E., Burns, M.S., & Griffin, P. (eds.) (1998). Preventing reading difficulties in young children. Washington, DC: National Academy 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* (100th Yearbook of the National Society for the Study of Education, Part II). Chicago: University of Chicago Press, 81–98.

Williams, K. T., Cassidy, J., Samuels, S. J. (2001). *GRADE Teacher's Scoring and Interpretive Manual Level 1*. San Antonio, TX: Pearson Education, Inc.

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 Started Kit*. Winston-Salem, NC: Child's Mind Publishing.

Yoon, K.S., Duncan, T., Lee, S. W.-Y., Scarloss, B., and Shapley, K. (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.

Appendix B. Tables and Figures

Table 1: Days and Per-Teacher Hours of PD Offered During the First Year of the Study

PD Activity	Colleague Classroom Teacher	Model Classroom Teacher	Principal	School-Based Coach
Summer Institute, days (hours)	3 days (24 hours)	3 days (24 hours)		3 days (24 hours)
Seminars, days (hours)	3 days (24 hours)	3 days (24 hours)		
Coaching, hours	50 hours	100 hours	7 hours	4 hours
Total PD hours	98 hours	148 hours	7 hours	20 hours
District-wide Meetings, #		6 meetings	6 meetings	

Table 2: Days and Per-Teacher Hours of PD Offered During the Second Year of the Study

PD Activity	Colleague Classroom Teacher	Model Classroom Teacher	Principal	School-Based Coach
Seminars, days (hours)	3 days (24 hours)	3 days (24 hours)		
Coaching, hours	25 hours	50 hours	7 hours	4 hours
Total PD hours	49 hours	74 hours	7 hours	4 hours
District-wide meetings, #, hours		9 meetings	6 meetings	

Table 3. Treatment and Control Group Contrast in Hours of Literacy-Related PD: Kindergarten and First Grade Second-Year Sample

	Treatment Group	Control Group	Estimated Difference	Standard Error of the Estimated Difference	Estimated Difference Effect Size	P-Value
Institutes or seminars (hours)	52.85	22.57	30.28	3.72	0.99	<.0001
Coaching (hours)	44.83	17.17	27.66	3.24	1.17	<.0001
Other PD (hours)	20.38	21.13	-0.75	5.93	-0.01	0.8992
Total PD (hours)	117.21	55.83	61.39	8.58	0.77	<.0001

Sample Size: 446 teachers (239 treatment and 207 control group teachers).

SOURCE: Spring Teacher Survey.

Table 4. Impact of the CLI Program on Teacher Instructional Practice: Second-Year Teacher Impact Analysis Sample

	Treatment Group	Control Group	Estimated Difference	Standard Error of the Estimated Difference	Estimated Difference Effect Size	P-Value
Classroom Environment ELLCO Scale	3.68	3.30	0.38	0.12	0.52	0.003*
Language and Literacy ELLCO Scale	3.28	2.82	0.46	0.11	0.68	<.001*

Sample Size: 130 teachers (65 treatment and 65 control group teachers).

SOURCE: ELLCO Teacher Observations.

NOTES: ELLCO scales are reported on a scale of 1 to 5, with 1 representing minimal/deficient evidence was observed in the classroom for the items comprising the scale, and 5 representing compelling/exemplary evidence observed in the classroom.

The analyses are based on a two-level model controlling for random assignment block, grade and school-average percent of LEP students.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for teachers in the treatment group as the basis for the adjustment.

Effect sizes were calculated using control group standard deviation for the teacher instructional practice impact analysis sample. The control group standard deviation was 0.73 for the *Classroom Environment ELLCO Score*, and 0.67 for the *Language and Literacy ELLCO Score.* P-values are based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

Table 5. Impact of the CLI Program on Kindergarten Literacy Achievement at the End of the First Year: Cohort 1 First-Year Student Impact Analysis Sample

	Treatment Group	Control Group	Estimated Difference	Standard Error of the Estimated Difference	Estimated Difference Effect Size	P-Value
Letter-Word Reading PAR Subscore	106.03	103.33	2.70	0.99	0.16	0.008*
Fluency PAR Subscore	101.61	100.38	1.23	0.73	0.09	0.097
Vocabulary PAR Subscore	89.42	90.14	–0.72	0.42	–0.05	0.093
Phonemic Awareness PAR Subscore	95.42	95.22	0.20	0.92	0.02	0.828
PAR Early Reading Skills Total Score	103.20	101.91	1.29	0.69	0.10	0.068

Sample Size: 4334 kindergarten cohort 1 students (including 2276 treatment and 2058 control group students).

SOURCE: Kindergarten Literacy Assessment - Predictive Assessment of Reading (PAR).

NOTES: The impact analyses for student literacy achievement were conducting using scale scores. The estimated impacts are based on a two-level model controlling for random assignment block, PAR pretest scores, and school-average percent of LEP students.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

Effect sizes were calculated using control group standard deviation for the student impact analysis sample. The control group standard deviation was 16.50 for the *letter-word reading PAR subscore*, 13.72 for the *fluency PAR subscore*, 12.46 for the *vocabulary PAR subscore*, 14.34 for the *phonemic awareness PAR subscore*, and 12.32 for the *PAR early reading skills total score*. P-values are based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

Table 6. Impact of the CLI Program on Kindergarten Literacy Achievement at the End of the Second Year: Cohort 2 Second-Year Student Impact Analysis Sample

	Treatment Group	Control Group	Estimated Difference	Standard Error of the Estimated Difference	Estimated Difference Effect Size	P-Value
Letter-Word Reading PAR Subscore	106.7	104.0	2.68	0.97	0.17	0.007*
Fluency PAR Subscore	102.3	101.1	1.15	0.79	0.09	0.151
Vocabulary PAR Subscore	88.9	88.4	0.50	0.44	0.04	0.262
Phonemic Awareness PAR Subscore	95.9	94.7	1.26	0.74	0.09	0.093
PAR Early Reading Skills Total Score	103.6	101.4	2.13	0.68	0.18	0.003*

Sample Size: 4321 kindergarten cohort 2 students (including 2309 treatment and 2012 control group students).

SOURCE: Kindergarten Literacy Assessment - Predictive Assessment of Reading (PAR).

NOTES: The impact analyses for student literacy achievement were conducting using scale scores. The estimated impacts are based on a two-level model controlling for random assignment block, PAR pretest scores, and school-average percent of LEP students.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

Effect sizes were calculated using control group standard deviation for the student impact analysis sample. The control group standard deviation was 15.75 for the *letter-word reading PAR subscore*, 13.40 for the *fluency PAR subscore*, 12.38 for the *vocabulary PAR subscore*, 14.29 for the *phonemic awareness PAR subscore*, and 11.98 for the *PAR early reading skills total score*. P-values are based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).

Table 7. Impact of the CLI Program on First Grade Literacy Achievement at the End of the Second Year: Cohort 1 Second-Year Student Impact Analysis Sample

	Treatment Group	Control Group	Estimated Difference	Standard Error of the Estimated Difference	Estimated Difference Effect Size	P-Value
Word Reading and Meaning GRADE Score	98.38	98.19	0.19	1.05	0.01	0.857
Comprehension GRADE Score	98.70	98.90	-0.20	1.12	-0.01	0.858
GRADE Total Score	95.96	96.09	-0.14	1.15	-0.01	0.907

Sample Size: 3937 first grade cohort 1 students (including 2112 treatment and 1825 control group students).

SOURCE: First Grade Literacy Assessment - Group Reading Assessment and Diagnostic Evaluation (GRADE).

NOTES: The impact analyses for student literacy achievement were conducting using scale scores. The estimated impacts are based on a two-level model controlling for random assignment block, PAR pretest scores, and school-average percent of LEP students.

The treatment and control columns display regression-adjusted mean outcomes for each group, using the mean covariate values for students in the treatment group as the basis for the adjustment.

Effect sizes were calculated using control group standard deviation for the student impact analysis sample. The control group standard deviation was 16.45 for the *word reading and meaning GRADE score*, 16.33 for the *Comprehension GRADE score*, and 16.96 for the *GRADE total score*. P-values are based on t-tests. Two-tailed statistical significance at the $p \leq .05$ level is indicated by an asterisk (*).