AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Children's Literacy Initiative

Final Report of the i3 Scale-up Study

SEPTEMBER 2020

Katie V. Drummond | Natalie Tucker-Bradway | Deeza-Mae Smith | Daniel Hubbard | John Meakin | Terry Salinger

MAKING RESEARCH **RELEVANT**

Children's Literacy Initiative

Final Report of the i3 Scale-up Study

SEPTEMBER 2020

Katie V. Drummond | Natalie Tucker-Bradway | Deeza-Mae Smith | Daniel Hubbard | John Meakin | Terry Salinger

AIR's mission is to generate and use rigorous evidence that contributes to a better, more equitable world.

AMERICAN INSTITUTES FOR RESEARCH®

1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3239
202.403.5000

www.air.org

Contents

Page

Tables

Page

Figures

Page

Abstract

This i3 scale-up study examined the implementation and effectiveness of a 3-year literacy intervention developed by the Children's Literacy Initiative (CLI). The study was a school-level cluster randomized controlled trial conducted in 55 elementary schools from four states. Implementation results showed a high level of fidelity of intervention implementation across the treatment schools in the four urban school districts that had not previously worked with CLI. Intent-to-treat analyses of observation data showed that the intervention's effects on classroom literacy environment and teacher practices were not statistically significant in Year 1. By Year 3, however, treatment teachers who had taught in intervention schools for all 3 years were rated significantly higher in both the quality of classroom environment and the quality of literacy instructional practices than their peers in control schools. Analyses of student achievement data indicated that the CLI intervention had no statistically significant impact on students' reading achievement in any of the 3 intervention years. This held true for the overall student sample and for grade-specific subsamples, based on the overall literacy test score and for literacy subtest scores. In addition, differential impact analyses revealed that the impact of the CLI intervention on students' literacy achievement did not differ significantly by the level of students' baseline achievement. Further evidence from exploratory analyses showed that the intervention had no statistically significant impact on the English proficiency of English learner students or on teachers' knowledge of beginning reading practices. Because of the high amount of teacher attrition, an additional exploratory analysis that restricted the sample to only students of teachers who were stable across all 3 years in both conditions. There were positive and statistically significant effects for students in the CLI intervention condition who had stable teachers (teachers who had greater opportunity for intervention exposure) and who entered in Grade 1 and progressed to Grade 3 during the course of the study.

Introduction

This study assessed the effectiveness of a structured literacy intervention developed by the Children's Literacy Initiative (CLI). The intervention is designed to improve classroom literacy environments and instruction as well as to raise the achievement of students in kindergarten through Grade 3. The CLI intervention provides teachers with training and coaching sessions, establishes mentor teachers to support fellow teachers, and involves school and district leaders in tracking students' literacy progress. Multifaceted reading and writing interventions—such as the CLI intervention—that involve a combination of materials and professional learning have become increasingly common as a way to supplement schools' existing reading programs, with the goal of improving student literacy outcomes. Literacy achievement has been a concern in the United States for many years, especially for the critical phase between kindergarten and Grade 3 during which students acquire foundational reading skills. A large body of research indicates that delivering reading instruction using evidence-based approaches in the early grades and intervening to support students who struggle can be effective for improving the reading skills for all students, including those with reading difficulties (e.g., Fletcher & Vaughn, 2009; Smith et al., 2016; Vellutino, Tunmer, Jaccard, & Chen, 2007).

Despite the rich evidence about effective approaches to reading instruction, many students do not become proficient with the foundational reading skills and are unprepared for more complex literacy tasks. As documented by Grade 4 students' scores on the most recent National Assessment of Educational Progress (National Center for Education Statistics, 2019), only 35% of the students scored at or above the proficient level, with another 31% scoring only at the basic level. Students who are Black, Hispanic, or American Indian have even lower proficiency levels, as do those who are ELs, students with disabilities, and students who are eligible for free or reduced-price lunch.

Although the importance of acquiring reading skills in a child's early school years is widely emphasized (Fiester, 2013; Hernandez, 2011), teachers may not always have the resources or specialized knowledge that they need to facilitate effective literacy instruction. Because establishing high-quality instructional practices for reading is nuanced and complex (Foorman Dombek, & Smith, 2016; Moats, 1994; Shanahan & Lonigan, 2010; Piasta, Connor, Fishman, & Morrison, 2009), it is hypothesized that professional development (PD) can be helpful for teachers to gain greater expertise in reading instruction. In a meta-analysis focused on the effects of teacher coaching, as a specific type of PD, Kraft, Blazar, and Hogan (2018) combined results across 60 causal research studies that investigated the effect of coaching programs. The synthesis found pooled effect sizes of .49 of a standard deviation (SD) for instructional practices and .18 SD for student achievement outcomes. The 35 studies that focused specifically on reading-focused coaching programs produced estimates of .51 SD for teachers' instructional

practice. The meta-analysis also showed that the quality and focus of the coaching programs were the most important factors related to the size of program effect, and these superseded the sheer number of coaching hours. This finding is consistent with those from other research that suggests teacher PD should strongly focus on content and involve active learning and collective participation of school staff (Garet et al., 2001).

In another meta-analysis focusing on PD interventions, Basma and Savage (2018) synthesized 17 studies with experimental or quasi-experimental designs that examined the impact of PD intended to improve reading instruction. These included approaches to improve teacher knowledge and teaching practice and strategies to help teachers better implement differentiated reading instruction. Of the 17 studies, 13 demonstrated positive effects on student literacy outcomes; the average effect size was nearly one fourth of a standard deviation (.23). However, the effects ranged widely from -.98 to .95 SD. Outcomes, such as teacher knowledge and classroom practice, may be more quickly influenced by a PD intervention compared to student achievement. For example, in an experimental study focused on a reading PD program provided over the course of one year (Garet et al., 2008), one group of teachers participated in a week-long intensive training institute, another participated in the institute along with classroom-embedded coaching, while a third group served as controls. Both the institute-only and the institute-plus-coaching groups had higher and statistically significant scores than control teachers on a test of teacher knowledge and on some aspects of instructional practice. However, differences in student reading achievement between each of the two treatment groups and the control group was not significant.

Prior to the current study, CLI's intervention was studied through a federal Investing in Innovation (i3) validation grant, in an experimental study involving kindergarten to Grade 2 classrooms in 78 schools in three districts. The study, conducted over 3 years, demonstrated positive effects on both teacher and student outcomes (Parkinson, Salinger, Meakin, & Smith, 2015). For teacher outcomes, a subset of 130 randomly selected kindergarten and Grade 1 teachers were observed 1.5 years into the study. Based on classroom observations using the Early Language and Literacy Classroom Observation tool (Smith, Brady, & Clark-Charelli, 2008), teachers in the treatment group had statistically significant and higher ratings than control group teachers on both (a) the classroom environment composite score (ES = .52) and (b) the language and literacy composite score (ES = .68). For student outcomes, among one cohort of 3,700 students who were followed for 3 years, the study showed that treatment students significantly outperformed control students on kindergarten prereading (ES = .12) and Grade 2 reading measures (ES = .13), with no statistically significant group differences in the Grade 1 reading outcome (ES = .01). Among a second cohort of 3,645 students who were followed for 2 years, there were statistically significant, positive treatment effects on students' kindergarten reading outcome (ES = .16), but again, the effect was not significant for Grade 1 (ES = .07).

In a previous quasi-experimental, matched comparison study of an early version of CLI's intervention, kindergarten and Grade 1 children in Philadelphia outperformed their counterparts at schools with similar achievement and demographics on district literacy skill assessments (OMG Center for Collaborative Learning, 2009). The evaluation also found that the CLI program facilitated positive relationships between and among teachers and administrators, which reflected CLI's aim to promote a strong atmosphere of professional learning in schools. Because literacy-focused PD can make a difference for teachers and students but also requires a large investment of time and money, it is important to identify effective literacy PD programs and investigate the ways in which they can be efficiently sustained and scaled. Given the strong evidence on the impact of the CLI intervention on teacher practice and promising evidence on the intervention's impact on student achievement found in prior research, the current study, funded through an i3 scale-up grant, aimed to further evaluate and validate the effectiveness of CLI's intervention when implemented on a large scale.

Overview of the CLI Intervention

CLI is a nonprofit organization that provides teachers and school- and district-level administrators with PD and resources with the goal of improving early literacy instruction, particularly for disadvantaged children. For this i3 scale-up study, CLI provided elementary schools with 3 years of the CLI intervention, which involves three components. The first component includes resources for K–3 teachers directly from CLI. These include training seminars, coaching, and facilitated grade-level meetings; in the first year, CLI also provides classrooms with books and materials aligned with the intervention model. A CLI-developed framework called Teacher's Effective Literacy Practice (TELP) outlines the scope and sequence for training and coaching. For the second component, CLI selects K–3 teachers from each participating school and trains them as instructional lead teachers (ILTs) to serve as mentors and develop a model classroom in which fellow teachers could observe best practices. Each ILT is assigned a specific topic of focus (e.g., readers' workshop) and receives additional coaching and training on that focus. For the third component, CLI holds school-level leadership team meetings and district-level meetings with principals and district administrators in an effort to involve school and district leaders with tracking progress of teacher and student literacy practice. According to the program theory of action, CLI expected that, once the program components were in place, intermediate outcomes of improved teacher knowledge and teacher practice should manifest first and, subsequently, longer term improvements in student achievement would manifest.

Table 1 illustrates the intended dosage for each of 12 indicators associated with three CLI intervention components, as implemented in this i3 scale-up study. It shows that a teacher who remained in a treatment school for all 3 program years was expected to participate in 56 hours

of seminar training, 105 hours of coaching, and 24 grade-level meetings. Teachers who were designated as ILTs and remained in treatment schools for 3 years were expected to receive 25 additional hours of coaching and six additional seminars on topics in their appointed area of expertise. To address teacher mobility, CLI included new teachers and long-term substitutes in their PD and identified and trained new ILTs each year, as needed.

Table 1. Design of the CLI Intervention, by Year of Implementation

Component	Indicator	Intended Dosage/Description		
		Year 1 (2016–2017)	Year 2 (2017–2018)	Year 3 (2018–2019)
Resources, Training, and Coaching for K–3 Teachers	1. Books and materials	Provided for all classrooms	n/a	n/a
	2. Seminars on early literacy instruction	3 days/ 24 hours	2 days/ 16 hours	2 days/ 16 hours
	3. Coaching for each teacher	35 hours	40 hours	30 hours
	4. Additional coaching at each school	10 hours	10 hours	10 hours
	5. Coverage of content areas on TELP checklist during coaching	≥ 3 areas	≥ 3 areas	≥ 3 areas
	6. CLI-facilitated grade-level meetings	8	8	8
Instructional Lead Teacher (ILT) Training and Coaching	7. Selection of ILTs	4 ILTs	6 ILTs	4 ILTs
	8. Coaching to each ILT	5 hours	10 hours	10 hours
	9. ILT seminars	2 seminars/ 12 hours	2 seminars/ 12 hours	2 seminars/ 12 hours
PD for School and District Leaders	10. CLI-facilitated school leadership meetings	4	4	4
	11. CLI-facilitated district meetings for principals	3	3	3
	12. CLI-facilitated district administrator meetings to review progress	2	2	2

n/a = not applicable

Study Design

This study assessed implementation of the CLI intervention and the impact of the intervention on classroom environment and teacher literacy practices and on student reading achievement. The intervention took place across 3 academic years: 2016–17 (Year 1), 2017–18 (Year 2), and 2018–19 (Year 3). We addressed the following primary research questions:

1. *Related to implementation*, to what extent is the CLI intervention implemented with fidelity to the proposed model? To what degree was there a service contrast between teachers receiving CLI services and teachers in the control group?
2. *Related to teacher practice*, what is the impact of the CLI intervention on classroom environment and literacy instructional practices after 1 and 3 years?
3. *Related to student achievement*, what is the impact of the CLI intervention on students' reading achievement after 1, 2, and 3 years? What are the impacts by grade-level subgroups?

In addition to the primary analyses, we conducted a series of exploratory analyses to answer additional research questions:

- What is the impact of the CLI intervention on particular subskills of reading and to what degree is there a differential impact for subgroups based on students' baseline achievement?
- What is the impact of the CLI intervention for students whose teachers had remained in the study schools for 3 years (and therefore treatment teachers had greater exposure to the CLI intervention)?
- What is the impact of the CLI intervention on EL students' English proficiency outcomes and on teachers' knowledge of beginning reading?

In the following sections, we provide more information about the study sample, measures, and analytic methods used to address the research questions.

Sample

CLI staff recruited four districts—Broward County Public Schools (Florida), Denver Public Schools (Colorado), Elizabeth Public Schools (New Jersey), and the Houston Independent School District (Texas)—to participate in the scale-up study. These districts were identified because they had a relatively high proportion of students who were below proficient on state ELA benchmarks and a high proportion of students from low-income families.

To recruit schools with conditions that would facilitate effective implementation of the intervention, CLI staff worked with superintendents and district staff to identify schools that had a minimum of two teachers per grade in Grades K–3. To minimize confounds, schools were deemed ineligible if they planned to implement other new, schoolwide literacy interventions during the study time period or if they had prior CLI intervention exposure. Finally, in order to serve EL students and because CLI previously found positive effects in schools with small percentages of EL students (i.e., 7% ELs in Parkinson et al., 2015), CLI targeted schools for this study with at least a 10% EL population.

In total, 55 schools were recruited to participate in the study and were randomized into treatment and control conditions. To minimize the chances of imbalance and improve the precision of impact estimates, randomization occurred within three blocks in each district for a total of 12 blocks. The blocks were created based on the percentage of Grade 3 students at each school that scored below, at, or above the proficient level on the state ELA achievement test in the prior year. Overall, 27 schools were assigned to the treatment group and 28 schools to the control group. As shown in Table 2, treatment schools and control schools were largely similar on demographic characteristics at baseline based on the full population of students in the schools. In addition, in a Year 1 fall survey treatment and control teachers reported similar background characteristics, including certification status (90% of treatment; 92% of control teachers had a standard teaching certificate or more advanced professional certificate), educational background (42% of treatment; 41% of control teachers had a master's degree or higher), average years of teaching experience (11 years for treatment; 13 years for control teachers), and similar rates of reading-focused specialization (33% of treatment; 29% of control teachers). In addition, treatment and control teachers reported that they had received similar amounts of literacy-focused training in the year prior to the start of the study (31 hours for treatment; 32 hours for control teachers)

Table 2. Baseline Characteristics of Study Schools, by Condition

District	Condition	Number of Schools	% Minority Students	% of Students Eligible for Free or Reduced Price Lunch	% of Grade 3 Students Proficient in Reading (2014–2015)	% of EL Students
Broward	Treatment	7	90%	88%	31%	24%
	Control	6	90%	91%	35%	29%
Denver	Treatment	7[b]	92%	86%	54%	59%
	Control	7	84%	88%	59%	53%
Elizabeth[a]	Treatment	6	93%	90%	17%	9%
	Control	8	92%	87%	18%	11%

District	Condition	Number of Schools	% Minority Students	% of Students Eligible for Free or Reduced Price Lunch	% of Grade 3 Students Proficient in Reading (2014–2015)	% of EL Students
Houston	Treatment	7	96%	84%	34%	48%
	Control	7	99%	94%	34%	55%

Source: Extant district records

[a]The number of schools differs between the treatment and control conditions in Elizabeth because of variation in the numbers of teachers and classes across schools and the desire to distribute treatment resources to similar numbers of students and teachers across districts.

[b]One of the seven treatment schools in Denver dropped out of the study prior to receiving the intervention.

Schools assigned to the treatment condition agreed to implement the CLI intervention for 3 years, while schools assigned to the control condition were to continue with business-as-usual literacy instruction and PD. Each treatment school received the CLI intervention free of charge. Each control school received a $2,000 stipend in Year 1 and a $1,500 stipend in both Years 2 and 3 to purchase book collections and/or other teaching resources.

School attrition was very low during the 3 years of the intervention. One school in the treatment group dropped out of the study (prior to any training in Year 1) because it decided it could not fulfill the requirements of the intervention. Across the 54 participating study schools, there were 18,151 K–3 students based on information provided in district rosters within the first month of school in fall 2016.[1] The study followed these students as they progressed across grade levels during the 3 years of the intervention. The design of the study meant that students in the treatment group had varying degrees of potential exposure to the K–3 teachers who had been trained by CLI and that those teachers could have increasing exposure to the CLI intervention each year. For example, a treatment student in Grade 3 at the beginning of the study would have up to one year of exposure to a CLI-trained teacher. Meanwhile, a treatment student in kindergarten or Grade 1 at the beginning of the study would have up to 3 years of exposure to CLI-trained teachers.

Across the 3 years, teacher attrition was relatively high. When the intervention began in fall 2016, the sample included 846 K–3 teachers (406 treatment and 440 control). By Year 2, 72% of those teachers remained in study schools (68% of the teachers in the Year 1 treatment group and 75% in the Year 1 control group). By Year 3, 51% of the Year 1 teacher pool remained in study schools (47% of treatment and 54% of control). Due to teacher attrition each year, many students in treatment group classrooms were taught by teachers who had been trained and

[1] Random assignment took place in spring 2016; therefore, it is possible some students and teachers entered the study schools after randomization (between spring and fall). However, random assignment results or schools' status as "CLI" or "Non-CLI" schools were not publicized, minimizing the possibility that teachers or students changed schools on the basis of treatment condition. No students or teachers were added to the impact samples after the research team's receipt of rosters in fall of Year 1.

coached by CLI for fewer than the full 3 years. Appendix A presents details on the size and flow of the sample.

Measures

To answer the primary research questions, the study team collected data to document (a) implementation fidelity of the CLI intervention and service contrast between teachers receiving CLI services and teachers in the control group; (b) the quality of teachers' classroom environment and literacy instruction, and (c) students' literacy achievement. In addition, we received administrative data from study districts annually, including rosters of teachers and students in each study school, and student demographic data, including student race and ethnicity, eligibility for free or reduced-price lunch, classification as EL students and their English language proficiency, and special education participation. For the exploratory research questions, we also collected extant data on EL students' English proficiency outcomes and measured teacher knowledge about beginning reading instruction (see Appendix B).

Measure of Implementation Fidelity

CLI staff collected various types of implementation data from study participants through coach logs, attendance records, and checklists. To develop a measure of implementation fidelity, the study team relied on the set of indicators outlined by CLI (detailed in Table 1) that would reflect the implementation fidelity for each of their intervention components:

Component 1: Resources, training, and coaching for K–3 teachers (six indicators in Year 1 and five indicators in Years 2 and 3)

Component 2: Instructional Lead Teacher training and coaching (three indicators for each year)

Component 3: PD for school and district leaders (three indicators for each year)

The study team, in collaboration with CLI, developed a scoring matrix to calculate the dosage for each indicator and the overall implementation fidelity of each key intervention component. First, each indicator was scored on a 3-point scale for the relevant participant. For instance, for teachers who were expected to attend 3 seminar training days, they would be assigned a "1" if one day was attended and "3" if all 3 days were attended. For indicators such as the number of coaching hours, points were assigned based on CLI's expectations for high participation–for example, a "3" if more than 75% of the expected number of coaching hours were delivered and a "1" if the number of coaching hours delivered fell at or below 50% of the expected number. Next, participant scores for a given indicator were averaged across participants in each school, and the cut points were applied to the school average to derive a school-level fidelity score for the indicator. Cut points for the school average indicator scores were established so that each school could be rated as *low*, *moderate*, or *high* on each indicator. Last, school-level fidelity

scores (1 = low; 2 = moderate; 3 = high) were averaged across all indicators associated with a given intervention component to derive a *sample-level fidelity rating* for the component. Appendix C presents the full implementation fidelity matrix.

Measures of Teachers' Professional Development

To examine the service contrast between teachers in treatment and control groups, the study team administered a teacher survey at multiple time points to collect information on teachers' experiences with PD. The baseline survey administered in the fall of Year 1 was designed to collect background information from teachers in terms of their education and certification levels, years of teaching experience, and classroom setting. The follow-up surveys administered in the spring of each intervention year collected information about the quantity of teachers' literacy PD activities, including training (e.g., courses, institutes, workshops), coaching, and peer learning opportunities, as well as the literacy topics of focus in these activities. The study team used multiple communication methods and monetary incentives ($15 gift cards) to encourage teacher participation in the surveys. Response rates averaged 61% across years, ranging between 50% and 66% at any given administration (see Appendix A for additional details).

Measure of Classroom Environment and Literacy Instructional Practices

To answer the research question related to the quality of teachers' classroom environment and literacy instructional practices, the study team observed a subsample of teachers during literacy instruction in Years 1 and 3. Observations were conducted using the Early Language and Literacy Classroom Observation (ELLCO) K–3 Research Tool (Smith et al., 2008). ELLCO is a validated observation tool containing 18 items across two broad subscales: *classroom environment* (seven items that document how well the classroom structure and curricular items support literacy) and *language and literacy instruction* (11 items that document instructional strategies for literacy domains, such as phonics, vocabulary, comprehension, writing, and discourse). Teachers received scores on the two subscales and an overall ELLCO score. Each ELLCO item is rated on a 5-point scale with "5" indicating a rating of "exemplary" (compelling evidence) and "1" indicating a rating of "deficient" (minimal evidence) of the construct.

Classroom observers attended training to learn how to observe a classroom related to each ELLCO item, how to write evidence statements, and how to assign scores. To be certified, observers watched and independently scored a set of practice videos and matched master ratings at 80% or above. Observers, who were blind to the study condition they were observing, scheduled observations in order to incorporate a teacher's full literacy block (up to 2 hours) on a day when typical instruction was scheduled. In the case of a shorter literacy block, observers observed a minimum of 90 minutes of instruction.

Observations were conducted once in the spring of Year 1 and again in the spring of Year 3. In Year 1, one teacher per grade (kindergarten through Grade 3) per school was randomly selected and a request was made to observe the teacher during ELA instruction (n = 216; 26% of all K–3 teachers). In total, 205 (95%) of the Year 1 observations were completed. In Year 3, among teachers who had remained in the study schools for 3 years, one teacher per grade per school was again randomly selected to participate in an observation. Because of attrition, there were cases in which no teachers remained for 3 years for certain treatment condition-by-grade-by-school combinations. In those cases, we selected a backup teacher who taught the same grade from another school in the same treatment condition within the same random assignment block. In total, 200 (93%) of 214 selected teachers were observed in Year 3. Teachers received a $50 gift card for their participation in each observation.

Student Literacy and Language Outcomes

The research questions about the impact of the CLI intervention on students' literacy outcomes were examined through two measures with two student subsamples.

Study-Administered GRADE Assessment Scores. As one measure of student reading achievement, the study team administered the Group Reading Assessment and Diagnostic Evaluation (GRADE; Pearson Education, 2014) at four timepoints: fall Year 1 (baseline), spring Year 1, spring Year 2, and spring Year 3. GRADE is an untimed, standardized, norm-referenced reading assessment that, in the early grades, measures four components of reading: (1) word reading, (2) word meaning, (3) sentence comprehension, and (4) passage comprehension. The kindergarten assessment measures early literacy skills of phonological awareness, print awareness, letter recognition, and sound-symbol correspondence. We used the GRADE total test score, a composite of the subtest scores, that is based on norms established for fall and spring for each grade level. GRADE has strong evidence of reliability and validity, including a high degree of internal consistency for GRADE composite and subtest scores for Grades K–5 (alphas between .95 and .99). The instrument also has high alternate-form reliability (.81–.94) and high test–retest reliability (.80). Concurrent validity studies of GRADE for Grades 1–6 indicate moderate to strong correlations between GRADE scores and scores on the Iowa Test of Basic Skills (.69–.90), as well as strong correlations between GRADE scores and scores on the Gates-MacGinitie Reading Tests (.86–.90; AGS Publishing, 2001). For this study, two forms of the GRADE were counterbalanced across test administrations. Using specifications provided in the GRADE testing manual, trained data collectors administered the untimed GRADE assessment using two sessions for each grade level in each school. The sessions took place in a separate classroom and lasted approximately 70–90 minutes in total, with breaks. Data collectors conducted make-up sessions, as needed, for students who were absent during one or both assessment sessions.

The GRADE was administered to a subset of randomly-selected students from each participating grade in each intervention year. In fall of Year 1, from the pool of eligible K–3 students in each school, we randomly selected 40 students per grade, distributed across all classrooms in a grade. The selected students remained the target sample for the GRADE assessment across the duration of the study (Tables A-1 and A-2 in Appendix A present detail on the GRADE sample flow and attrition). Because delivering the GRADE assessment in languages other than English and delivering extensive special education accommodations were beyond the scope of the study, two criteria were applied to student selection. First, students needed to have sufficient English proficiency to take the GRADE assessment in English. Students' English proficiency level was based on outcomes from state-specified English proficiency assessments (See Appendix A, Table A-3), using the same criterion the district uses to identify students' language of assessment. Students who needed special education or 504 accommodations were included in the sample if accommodations fell into the following categories: directions read aloud; repeat, clarify, or reword directions; sitting near or facing the tester; extended time; and frequent breaks, opportunities to stand or move during the assessment session.

Extant State English Language Arts (ELA) Assessment Scores. As a second measure of students' literacy outcomes, the study team obtained students' reading scores on end-of-year standardized state ELA assessments for the subset of students who took these tests. These included the Florida Standards Assessments (FSA) for students in Broward; the Colorado Measures of Academic Success (CMAS) for Denver, and the State of Texas Assessments of Academic Readiness (STAAR) for Houston. Elizabeth used the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment in Years 1 and 2 and changed to the New Jersey Student Learning Assessment (NJSLA) in Year 3. Although the ELA assessments vary across states, they commonly measure the following reading skills: (a) understanding of main ideas and key details in literary and informational texts and (b) determining or identifying meaning of words. Students begin taking state ELA assessments in Grade 3 and continue taking the assessments as they progress through elementary school. In the study, only Grade 3 students had state assessment data in Year 1; these students were followed in Years 2 and 3 and included in state assessment analyses as Grades 4 and 5 students. In Years 2 and 3, we also added grade-level cohorts of students who had been in Grade 1 or 2 at the start of the study but took the state assessments as they entered Grade 3. To put the standardized ELA reading test data from different states on a common scale, we standardized students' scores based on the state means and standard deviations within year and grade.

Impact Analytic Methods

For the impact analyses, we used an intent-to-treat (ITT) framework to estimate the effects of the CLI intervention on teacher and student outcomes.

Teacher Observations Model

To answer the research question about the intervention's effects on teacher outcomes (i.e., the quality of classroom environment and literacy instruction) based on classroom observations, the study used a two-level hierarchical linear model (HLM) with teachers nested within schools. The model includes block fixed effects and uses a pooled-sample approach combining data from all target grades across all four districts in the study sample. The variables in this model were not centered. The model is specified as follows:

Level 1 (Teachers): $Y_{jk} = \beta_{0k} + \beta_{1k}\,(Grade)_k + \varepsilon_{jk}$

where

- Y_{jk} is the outcome (either the ELLCO overall score or a subscale score) for teacher *j* in school *k;*
- β_{0k} is the intercept, which is the mean outcome across all teachers in school *k*;
- β_{1k} is the relationship between the teacher's grade of instruction and the classroom observation measure for a teacher in school *k;*
- $(Grade)_k$ is a vector of grade of instruction indicators equal to 1 for kindergarten, Grade 1, Grade 2, and Grade 3; and
- ε_{jk} is a random error associated with teacher *j* in school *k,* assumed to be independent and identically distributed.

Level 2 (Schools): $\beta_{0k} = \sum_{b=1}^{B} \gamma_{00b}(Block)_{bk} + \sum_{w=1}^{W} \gamma_{01w}(Treat \times District_w)_k + r_{0k}$

where

- γ_{00b} is the average outcome across teachers in control schools in random assignment block *b*;
- $(Block)_{bk}$ equals 1 if school *k* is in block *b* and 0 otherwise;
- γ_{01w} is the treatment effect in district *w* on the teacher outcome;
- $District_{wk}$ equals 1 if school *k* is in district *w* or 0 otherwise;
- $Treat_k$ equals 1 if school *k* was assigned to receive the treatment and 0 otherwise; and
- r_{0k} is a random error associated with school *k,* assumed to be independent and identically distributed.

The average of the estimated γ_{01w} coefficients for the four districts, weighted by the number of treatment schools in each district, is the estimated intervention effect on the given teacher outcome for the average treatment school in the study sample. To assess whether the average

treatment effect differs from zero, we conducted a two-tailed *t*-test. We computed the average treatment effect on each teacher outcome as an ES, based on the pooled within-group standard deviation for the given year.

Student Outcome Model

To answer the research questions about student achievement, we used a three-level HLM, in which students are nested within classrooms and classrooms are nested within schools. This model included block fixed effects and used a pooled-sample approach combining data from all relevant grades across the four study districts. This model was estimated separately for each student outcome, including the GRADE overall scale score and subscale scores, state standardized ELA assessment scores (normalized to have mean zero and standard deviation 1 for all states, grades, and years), and measures of English language proficiency. All the variables in this model were uncentered. For all continuous outcomes, the model is specified as follows:

Level 1 (Students): $Y_{ijk} = \beta_{0jk} + \beta_{1jk}(baseline)_{ijk} + \sum_{p=2}^{P} \beta_{pjk} X_{pijk} + \varepsilon_{ijk}$

Where:

- Y_{ijk} is the value of the student-level outcome variable *Y* for student *i* taught by teacher *j* in school *k*;
- $(baseline)_{ijk}$ is the baseline GRADE score for student *i* taught by teacher *j* in school *k*;
- X_{pijk} is a vector of student characteristics, including gender, age, race/ethnicity, disability status, and EL status, along with grade indicators and an indicator for whether the student is missing baseline test score[2];
- β_{0jk} is the intercept, which is the mean of the outcome variable for all students taught by teacher *j* in school *k*, adjusted for student baseline outcome measure and other characteristics;
- β_{1jk} is the relationship between the baseline measure and the outcome for students taught by teacher *j* in school *k*;
- β_{pjk} is the relationship between a given student characteristic X_p and the outcome for students taught by teacher *j* in school *k*; and
- ε_{ijk} is a random error associated with student *i* taught by teacher *j* in school *k*, assumed to be independent and identically distributed.

[2] Students with missing baseline test scores have the baseline score variable set equal to the mean non-missing score within a school and grade (El-Masri & Fox-Wasylyshyn; 2005).

Level 2 (Teachers): $\beta_{0jk} = \delta_{00k} + r_{0jk}$

$$\beta_{1jk} = \delta_{10k}$$

$$\beta_{pjk} = \delta_{p0k}, p = \{2, \ldots, P\}$$

Where:

- δ_{00k} is the average outcome for all students taught by study teachers in school *k,* adjusted for student baseline scores and other characteristics*;*
- δ_{10k} is the average relationship between the baseline measure and the student outcome Y_{ijk} in school *k;*
- δ_{p0k} is the average relationship between the student characteristic indexed by *p* {2, …, *P*} and the student outcome Y_{ijk} in school *k; and*
- r_{0jk} is a random error associated with teacher *j* in school *k*.

Level 3 (Schools): $\delta_{00k} = \sum_{b=1}^{B} \gamma_{000b}(Block)_{bk} + \sum_{w=1}^{W} \gamma_{001w}(Treatment \times District_w)_k + u_{00k}$

$$\delta_{10k} = \gamma_{100}$$

$$\delta_{p0k} = \gamma_{p00}, p = \{2, \ldots, P\}$$

Where:

- $Block_{bk}$ is an indicator for whether school *k* is in block *b*;
- $District_{wk}$ equals 1 if school *k* is in district *w* and 0 otherwise;
- $Treatment_k$ equals 1 if school *k* was assigned to the treatment group and 0 otherwise;
- γ_{000b} is the average student outcome in control schools in block *b,* adjusted for student baseline outcome measure and other characteristics;
- γ_{001w} is the treatment effect in district *w* on student achievement;
- γ_{100} is the average relationship between the baseline measure and the outcome Y_{ijk} across all schools;
- γ_{p00} is the average relationship between the student characteristic with index *p* {2, …, P} and the outcome Y_{ijk} across all schools; and
- u_{00k} is a random error associated with school *k.*

The average treatment effect of the CLI intervention on a student outcome across all study districts is the linear combination of all values of γ_{001w}, weighted by the number of treatment schools in each district. To assess whether the average treatment effect differs from zero, we conducted a two-tailed *t*-test. We also computed the average treatment effect as an ES, based on the pooled within-group standard deviation of the outcome measure.

Findings

In this section, we present results for Research Question 1 related to implementation. We provide data on teachers' reported literacy-focused PD across the 3 years to consider the service contrast between treatment and control groups. Next, we report results on fidelity of implementation for the CLI intervention.

Implementation

In surveys, completed in the spring of each year, teachers were asked to report the number of hours of literacy-related PD they had received that school year. Each year, treatment teachers reported approximately 65% more hours per year of training (e.g., institutes, workshops), on average, compared to control teachers. Although there was a significant amount of variation among teachers' self-reported hours, the difference between groups was statistically significant each year, confirming a contrast in literacy-related training between treatment and control groups (Table 3).

Table 3. Average Number of Literacy-Focused Training Hours, by Year and Study Condition

Year	Treatment		Control	
	Number of Teachers	Adjusted Mean (SD)	Number of Teachers	Adjusted Mean (SD)
Year 1	198	57.0** (49.5)	216	36.4 (46.4)
Year 2	302	61.5** (50.5)	301	39.0 (44.0)
Year 3	269	61.7** (45.3)	226	40.0 (41.8)

Source: Study-administered survey of teachers.

Notes: Treatment means are adjusted for block and grade fixed effects.

**Difference between treatment and control was statistically significant at $p < 0.01$.

Similarly, treatment teachers reported more hours of literacy-focused coaching compared to control teachers. Again, the difference between groups was statistically significant each year (Table 4).

Table 4. Average Number of Literacy-Focused Coaching hours, by Year and Study Condition

Year	Treatment		Control	
	Number of Teachers	Adjusted Mean (SD)	Number of Teachers	Adjusted Mean (SD)
Year 1	198	32.3** (17.2)	216	6.9 (15.1)
Year 2	289	36.5** (20.8)	289	7.2 (13.7)
Year 3	266	32.6** (17.8)	230	6.7 (13.5)

Source: Study-administered survey of teachers.

Notes: Treatment means are adjusted for block and grade fixed effects.

**Difference between treatment and control was statistically significant at $p < 0.01$.

On the surveys, teachers indicated the degree to which various literacy topic areas were emphasized in the PD they received. Topics that the largest percentage of treatment teachers in Year 1 rated as being a major emphasis of training included read-alouds (72%), classroom literacy environment (56%), reading comprehension (54%), classroom management (52%), and reading workshops (49%). According to treatment teacher surveys in subsequent years, classroom management and literacy environment were emphasized less, as topics shifted to guided reading (48% of Year 2 and 60% in Year 3), small-group instruction (45% in Year 2 and 49% in Year 3), and writing (53% in Year 3). In contrast, control teachers rated the following topics as being a major emphasis of training in Year 1: interpretation of assessment data (58%), comprehension (57%), guided reading (50%), and differentiated instruction (47%). In Years 2 and 3, more control teachers also rated read-alouds (37% in Year 2 and 42% in Year 3) and writing (34% in Year 2 and 41% in Year 3) as being a major emphasis of training. Appendix D presents the full set of literacy topics and teacher ratings for both trainings and coaching each year.

In Years 2 and 3, teachers responded to a survey question asking them to rate five qualitative aspects of the literacy coaching they received in the last year. Teachers in both conditions held favorable perceptions, with a strong majority (greater than or equal to 70%) agreeing or strongly agreeing with the positive attributes of the coaching they received. In Year 1, CLI treatment teachers were more likely than control teachers to report they had a good relationship with their coach and felt comfortable receiving feedback from their coach. By Year 3, four of the five coaching attributes were rated more highly by CLI treatment teachers with statistically significant differences compared to the control group (Table 5).

Table 5. Percentage of Teachers Who Agreed and Strongly Agreed With Statements About Features of the Literacy Coaching They Received, by Year and Study Condition

Coaching Features	Treatment Teachers		Control Teachers	
	Year 2 (N = 270)	Year 3 (N = 272)	Year 2 (N = 132)	Year 3 (N = 110)
Coach promotes a positive learning environment.	90%	88%**	83%	76%
Coach has positive impact on the teachers' PD.	84%	89%*	85%	79%
Teachers have good relationship with the coach.	90%*	94%**	80%	76%
Teachers are comfortable receiving feedback from the coach.	93%*	94%**	83%	80%

Coaching Features	Treatment Teachers		Control Teachers	
	Year 2 (N = 270)	Year 3 (N = 272)	Year 2 (N = 132)	Year 3 (N = 110)
Coach's feedback was helpful.	85%	70%	85%	79%

Source: Study-administered survey of teachers.

Notes: Treatment means are adjusted for block and grade fixed effects.

*Difference between treatment and control was significant at $p < 0.05$; **difference between treatment and control was significant at $p < 0.01$.

Implementation Fidelity

Overall, the fidelity of implementation of the CLI intervention was rated high across the three components of the intervention over the 3 years of implementation. The first component, reflecting the resources, training, and coaching from CLI to teachers, had a "high" fidelity rating in Year 1 through Year 3. The second component, relating to the ILTs, the group of teachers who CLI selected and trained to mentor fellow teachers, was somewhat less consistent in terms of implementation fidelity across years, with a fidelity rating of "high" in both Years 1 and 2 and "moderate" in Year 3. The third component, PD for school and district leaders, was rated "high" on implementation fidelity in Years 1 through 3 (see Table 6). The combination of high attrition and variation in implementation resulted in only 19% of the treatment teachers who began the intervention in Year 1 receiving high-dosage levels of the CLI training and coaching consistently across the 3 years. Note that CLI delivered make-up PD to Grades K–3 teachers who entered treatment schools anew in Years 2 and 3, but these teachers received less than the full dosage of the intervention.

Table 6. Average Fidelity Ratings Across Treatment Schools, by Component and Year of Study

Component	Year 1 (2016–2017)			Year 2 (2017–2018)			Year 3 (2018–2019)		
	School Range	Average	Level	School Range	Average	Level	School Range	Average	Level
Resources, training, and coaching for teachers	2.67 to 3.00	**2.96**	**High**	2.40 to 3.00	**2.85**	**High**	2.40 to 3.00	**2.92**	**High**
Appointing and training and coaching for ILTs	2.33 to 3.00	**2.88**	**High**	1.67 to 3.00	**2.56**	**High**	1.00 to 3.00	**2.45**	**Moderate**
PD for school and district leaders	2.33 to 3.00	**2.83**	**High**	2.33 to 3.00	**2.87**	**High**	2.00 to 2.33	**2.65**	**High**
Overall:	n/a	**2.89**	**High**	n/a	**2.76**	**High**	n/a	**2.68**	**High**

Source: CLI implementation data

Impact

Next, we present results for primary Research Questions 2 and 3 on the impact of the CLI intervention on classroom environment and literacy instructional practices and on students' reading achievement. Appendix B presents findings for the exploratory analyses related to the impact of the CLI intervention for reading subskills, students with stable teachers, English proficiency outcomes, and teacher knowledge.

Impact on Classroom Environment and Literacy Instructional Practices

To examine the impact of the CLI intervention on classroom environment and teachers' literacy instructional practices after 1 and 3 years of participation, we used data collected with the ELLCO observation tool. In Year 1, differences between the two study groups were not statistically significant for either ELLCO subscale or the ELLCO total score. By Year 3, there were statistically significant differences between the treatment and control groups for the classroom environment subscale (ES= 0.49; $p < .01$), the language and literacy subscale (ES= 0.39; $p < .01$), and the combined score (see Table 7).

Table 7. Effects of the CLI Intervention on Classroom Environment and Teacher Literacy Practices, by Year

ELLCO Component	Statistics	Year 1	Year 3
Classroom Environment	ES	0.149	0.487
	p-Value	0.293	0.003**
Language and Literacy	ES	0.073	0.387
	p-Value	0.600	0.008**
Combined Scales	ES	0.112	0.468
	p-Value	0.426	0.002**
n/a	Sample Size	205	200
	Treatment	99	105
	Control	106	95

Source: Study-administered ELLCO teacher observations

Notes. The Year 1 sample contained 54 schools: 26 in the CLI group and 28 in the control group; the Year 3 sample contained 50 schools: 24 in the CLI group and 26 in the control group. *P*-values are based on two-tailed *t* tests; *p<.05; **p<.01.

Impact on Student Performance on Study-Administered GRADE Assessment

In the fall of Year 1, 8,379 students were randomly selected to take the GRADE exam. Of these students, 29% (2,455 students) were ELs, and 7% (581 students) had special education needs and received appropriate accommodations. Parental consent to participate in the student

assessment was received for 77% (6,491 students) of the subsample. The subsample consisted of two sets of students: (1) students who were in kindergarten and Grade 1 when the study began and took the GRADE assessment all 3 years of the study and (2) students who were in Grades 2 or 3 when the study began and who aged out of the analytic sample when they reached Grade 4. For both sets, attrition occurred when a student moved out of a study school or was unable to be tested, either due to lack of parental consent or unavailability on testing days (Details on sample flow in Appendix A, Table A-1).

Table 8 presents the effects of the CLI intervention on students' total GRADE scores for each of the 3 years based on data pooled across grades. ESs were low (ranging between -.01 and .04) and the difference between the treatment and control groups in students' GRADE scores was not statistically significant in any of the 3 intervention years.[3]

Table 8. Effects of the CLI Intervention on Total GRADE Scores For All Grades, by Year

	Year 1 (Grades K–3)	Year 2 (Grades 1–3)	Year 3 (Grades 2–3)
ES	-0.005	0.010	0.041
p-Value	0.894	0.861	0.580
Sample size	5,659	3,409	1,949
Treatment	2,792	1,664	943
Control	2,867	1,745	1,006

Source: Study-administered GRADE assessment

Notes. Sample contained 54 schools; 26 in the CLI group and 28 in the control group in each year. *P*-values are based on two-tailed *t* tests; *p<.05; **p<.01.

In addition to assessing the overall effect of CLI on students' reading achievement across grades, we also explored CLI effects by grade level. Table 9 presents results separately by the grade level of students when they entered the study in Year 1. In general, ESs were low and differences between treatment and control groups were not statistically significant. However, the subsample of students who were in Grade 1 at the start of the study and in Grade 3 at its conclusion had noticeably larger ESs in Years 2 and 3 (ES = .15 and .14, respectively), compared with students who started the study in other grades (ES ranging between -.10 and .03). As detailed in Appendix B, among this same group, ESs were even higher and statistically significant, in exploratory analyses that restricted the sample to only students of teachers who were stable across all 3 years in both conditions (including intervention teachers who had greater opportunity for intervention exposure).

[3] Appendix E presents information on the baseline equivalence of the analysis samples.

Table 9. Effects of the CLI Intervention on Overall GRADE Scores, by Grade and Year

Entry Grade[a]	Statistics	Year 1	Year 2	Year 3
Kindergarten	ES	-0.053	-0.101	-0.050
	p-value	0.505	0.280	0.582
	Sample size	1,389	1,161	1,010
	Treatment	695	581	497
	Control	694	580	513
Grade 1	ES	0.096	0.148	0.144
	p-value	0.104	0.073	0.092
	Sample size	1,378	1,069	937
	Treatment	660	514	444
	Control	718	555	493
Grade 2	ES	0.030	0.006	n/a
	p-value	0.630	0.944	n/a
	Sample size	1,430	1,178	n/a
	Treatment	700	568	n/a
	Control	730	610	n/a
Grade 3	ES	-0.081	n/a	n/a
	p-value	0.250	n/a	n/a
	Sample size	1,462	n/a	n/a
	Treatment	737	n/a	n/a
	Control	725	n/a	n/a

Source: Study-administered GRADE assessment

Notes. Sample contained 54 schools; 26 in the CLI group and 28 in the control group in each year. n/a = not applicable, aged out of testing pool. *P*-values are based on two-tailed *t* tests; *p<.05; **p<.01.

[a]*Entry grade* refers to the grade level of the students in Year 1 of the study.

Impact on Student Performance on State ELA Assessments

Results from the analyses of students' performance on state end-of-year standardized ELA assessments are presented in Table 10. Because students began taking these assessments in Grade 3, the Year 1 analysis included only students from Grade 3 in Year 1, who were followed in subsequent years (Grades 4 and 5. The Year 2 and Year 3 analyses included students from additional entry grades, as they entered Grade 3 and took the standardized state assessments. Information on the baseline equivalence of the analytic samples is presented in Appendix E.

Table 10. Effects of the CLI Intervention on Student Performance on State Standardized ELA Assessments, by Grade and Year

Entry Grade[a]	Statistics	Year 1	Year 2	Year 3
Grade 1	ES	n/a	n/a	0.118
	p-value			0.102
	Sample size			2,663
	Treatment			1,201
	Control			1,462
Grade 2	ES	n/a	-0.027	-0.033
	p-value		0.708	0.698
	Sample size		3,281	2,971
	Treatment		1,457	1,342
	Control		1,824	1,629
Grade 3	ES	-0.096	0.039	-0.069
	p-value	0.347	0.571	0.536
	Sample size	4,122	3,233	2,925
	Treatment	1,978	1,524	1,405
	Control	2,144	1,709	1,520
All students	ES	n/a	0.007	0.018
	p-value		0.905	0.772
	Sample size		6,514	8,559
	Treatment		2,981	3,948
	Control		3,533	4,611

Note. n/a = not applicable, due to grade levels when test is administered.

[a]*Entry grade* refers to the grade level of students when they entered the study in Year 1.

As shown in Table 10, the CLI intervention did not have a statistically significant effect on students' performance on state ELA assessments for the overall study sample or for subsamples defined by entry grade in any of the 3 intervention years. However, in Year 3, the ES (0.12) for the subsample of students who entered the study in Grade 1 is noticeably larger than that for students who started the study in other grades. This finding is consistent with a similar pattern in the GRADE assessment results.

Summary and Discussion

Summary of Findings

This study involved a large effort by CLI, researchers, and participating districts to assess the implementation and impacts of the multi-year, multi-component CLI literacy intervention in 54 elementary schools within four districts. Implementation findings indicate that treatment schools achieved high levels of fidelity of implementation overall. The first and third intervention components, related to the teacher-level and leadership-level implementation, respectively, received a high fidelity rating in each year of the intervention. The remaining component, related to ILTs, was rated high on implementation fidelity in Years 1 and 2 and moderate in Year 3.

This study showed that effects on teachers' literacy teaching strategies and literacy environment did not emerge after one year of the intervention, but after 3 years, treatment teachers who had taught in intervention schools for all 3 years were rated significantly higher in both the quality of classroom environment and the quality of literacy instructional practices than their peers in control schools. Despite the relatively high levels of fidelity of implementation in treatment schools each year and the positive impact of the CLI intervention on teacher outcomes at the end of Year 3, the intervention did not produce a statistically significant impact for any of the reading achievement outcomes in any of the 3 intervention years. This finding held true for the overall student sample and for grade-specific subsamples based on the overall literacy test score. The subsample of students who were in Grade 1 at the start of the study and in Grade 3 at its conclusion had larger ESs in Years 2 and 3 compared to other grade levels. Exploratory analyses showed no consistent pattern of statistically significant impacts for particular literacy subskills. In addition, analyses conducted on the basis of students' baseline literacy achievement showed results did not differ significantly by level of achievement. Further evidence from exploratory analyses showed that the intervention had no statistically significant impact on the reading domain of English proficiency assessments for EL students or on teachers' knowledge of beginning reading practices. In exploratory analyses that restricted the sample to only students of teachers who were stable across all 3 years in both conditions (including intervention teachers who had greater opportunity for intervention exposure), there were higher ESs that were statistically significant.

Implications and Directions for Future Research

This study adds an important piece to the evidence base for CLI's multi-component literacy intervention when the program is implemented at scale for multiple years. This study is especially relevant, given that many studies of reading interventions assess student and teacher outcomes after only 1 year of implementing a new program. This study shows that improvement in literacy teaching strategies may take multiple years to translate into improved

classroom practice. These effect sizes for teacher practice were similar to the previous 3-year validation RCT study of CLI (Parkinson et al., 2015) and to findings from other syntheses of literacy-oriented teacher PD (e.g., Kraft, Blazar, & Hogan, 2018). However, in the current study, the positive impact on teacher practice outcomes did not translate to positive impact on students' literacy achievement for the overall sample, as they had in the previous CLI study that showed significant positive impacts on student achievement outcomes in each of 2 intervention years and for two of the three early grade levels (Parkinson et al., 2015). In their meta-analysis of 60 coaching studies (many of which also involved additional training and seminar days, like CLI's program), Kraft et al. (2018) found that large improvements in instructional quality (1 SD) were typically required for associated changes in student achievement (.21).

There are several notable differences between this and the earlier large study of the CLI intervention, which may at least partly explain the different findings from the two studies. In the previous validation study, the intended dosage of literacy PD was higher: teachers were provided with 98 hours of PD in Year 1 across training and coaching activities; they were provided with another 49 hours in Year 2 (total of 147 hours after 2 years) and another 39 hours in Year 3 (total of 186 hours after 3 years). For the current scale-up study, the intended dosage was substantially lower in Year 1 and the same for Years 2 and 3: 59 hours in Year 1, another 49 hours in Year 2 (total of 108 hours after 2 years), and another 39 hours in Year 3 (total of 147 hours after 3 years). Yet, these amounts still equal a relatively large number of hours when compared to other teacher PD-focused interventions (Kraft et al., 2018; Matsumura et al., 2010) and in consideration of research that indicates it may take only 30 hours of PD to make a meaningful difference in instruction and student outcomes (Basma & Savage, 2017; Guskey & Yoon, 2009).

Another difference between the current scale-up study and the previous validity study is that schools in the previous validation study had a lower percentage of EL students than the current study. For this study, CLI purposely recruited schools with relatively higher EL rates, given that positive findings in the validation study occurred with a sample in which treatment schools had a higher proportion of ELs.[4] However, differences in the sample composition between the current study and the previous validation may have led to different study results.

Participant attrition could also have influenced study results in that attrition reduced teachers' and students' total exposure to the intervention. Although the services that CLI directly delivered to teachers were at a high level in each year of the intervention, by Year 3, only 49% of the original treatment teachers were still in the study schools overall, and only 19% of the original treatment teachers received high ratings of PD fidelity across all 3 intervention years. The high

[4]In another previous study of an earlier version of the CLI intervention (OMG, 2009), results for the study's Latino ELL kindergarten and Grade 1 students were more varied compared to non-ELLs.

rate of attrition over time also might have diminished the effect of CLI's efforts to establish a corps of ILTs, who could become experts and model the assigned focal teaching and from whom other teachers could learn. This level of teacher turnover may not be unusual for many urban schools. For instance, previous studies in districts of varying sizes have reported school-level teacher attrition between 45–63% over 5 years (e.g., Marinell & Coca, 2013; Papay, Bacher-Hicks, Page, & Marinell, 2017), and it has also been noted that schools serving underachieving students are the most likely to experience teacher instability (Atteberry, Loeb, & Wyckoff, 2017). However, in CLI's previous validation study, 76–87% of teachers were still in the same schools after 2 years, and 66% of teachers were in the same schools after 3 years. In another multi-year RCT study of a literacy coaching program that experienced high yearly teacher mobility (nearly 50%), Matsumara and colleagues (2010) suggest that teacher turnover in school-level reforms can result in school administrators and instructional coaches diverting focus from deepening new instructional practices to bringing new teachers up to speed. Yet, they attributed positive impacts for teacher practice and student achievement in their study to the fact that coaching activities and a new professional culture in treatment schools became well established in the first year, creating a foundation and model for teachers new to the schools.

In the current study, it is possible that the level of teacher PD for control schools was high enough to diminish potential effects of the CLI intervention. Although there was a contrast between treatment and control teachers' reported hours of literacy-related PD, control teachers still reported receiving almost a full week of literacy training (compared to a week and a half reported by treatment teachers) and almost a full day of coaching support (compared to 3–4 days for treatment teachers). We did not have a detailed measure to document the quality and precise content or approach of the literacy PD that control teachers received, but it is possible that the "business-as-usual" training and coaching focused on factors similar to those presented by CLI.

Future research and development with the CLI intervention could benefit from more targeted recruitment so that samples of districts and schools would allow for further investigation of how to best address the reality of teacher and student mobility in the context of an intervention that may need multiple years and cumulative dosage to make a difference. Research on systems change, including adoption of interventions in educational settings (Fixsen et al., 2005; Joyce & Showers, 2002), has documented the time-intensive and iterative nature of multiple stages involved with implementing new programs. Fixsen and colleagues (2005) describe this as a process in which practitioners transform from (a) gaining formal knowledge on new program approaches to (b) integrating these strategies into their existing skill sets. Full implementation may take up to 2 to 4 years, which stands in contrast to demands in the field for high accountability for immediate student achievement results (Blasé, Fixsen, Sims, & Ward, 2014; Gill et al., 2005).

It is important to continue investigating ways through which the CLI intervention model can be further optimized for delivery in schools in which teachers (and by association, their students) may only experience 1–2 years of the intervention. In the current study, we were limited to conducting just one observation per teacher at two time points. Future studies could benefit by conducting additional classroom observations in order to more precisely track the progression of teachers' development and to determine the timeframe in which the CLI intervention can affect changes in literacy teaching practices. Additional studies could add more qualitative measures, such as interviews or focus groups, to better understand teacher and administrator perceptions and use of CLI strategies while CLI is involved at the school and also after CLI's direct involvement ends. Studies could also include data collection from coaches to gain their perceptions of their training for and challenges of their role and of teachers' needs. Although teachers who completed the survey rated the CLI coaching very positively, we do not know from the current data whether there was some content of the CLI training that was more difficult or took more time to understand or implement. Finally, future studies would benefit from more detailed measures that capture the service contrast between CLI schools and business-as-usual schools.

References

Atteberry, A., Loeb, S., & Wyckoff, J. (2017). Teacher churning: Reassignment rates and implications for student achievement. *Educational Evaluation and Policy Analysis, 39*(1), 3–30.

Basma, B., & Savage, R. (2017). Teacher professional development and student literacy growth: A systematic review and meta-analysis. *Educational Psychology Review, 30*, 457–481.

Blasé, K. A., Fixsen, D. L., Sims, B. J., & Ward, C. S. (2015). *Implementation science: Changing hearts, minds, behavior, and systems to improve educational outcomes.* Oakland, CA: The Wing Institute.

El-Masri, M., & Fox-Waysylyshn, S. (2005). Missing data: An introductory conceptual overview for the novice researcher. *Canadian Journal of Nursing Research, 37*(4), 156–171.

Fiester, L. (2013). *Early warning confirmed: A research update on third grade reading.* Baltimore, MD: The Annie E. Casey Foundation.

Fixsen, D. L., Naoom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation Research: A synthesis of the literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30–37.

Foorman, B. R., Dombek, J., & Smith, K. (2016). Seven elements important to successful implementation of early literacy intervention. *New Directions for Child and Adolescent Development, 154*, 49–65.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., … Sztejnberg, L. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

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.

Gill, B. P., Hamilton, L. A., Lockwood, J. R., Marsh, J. A., Zimmer, R. W., Hill, D., & Prebish, S. (2005). *Inspiration, perspiration, and time: Operations and achievement in Edison schools.* Santa Monica, CA: Rand Corporation.

Guskey, T. R., & Yoon, K. S. (2009). What works in professional development? *Phi Delta Kappan, 90*(7), 495–500.

Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation.* New York, NY: The Annie E. Casey Foundation.

Joyce, B., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal Evidence. *Review of Educational Research, 88*(4), 547–588.

Marinell, W. H., & Coca, V. M. (2013). *Who stays and who leaves? Findings from a three-part study of teacher turnover in NYC middle schools.* New York, NY: The Research Alliance for New York City Schools.

Matsumura, L. C., Garnier, H. E., Correnti, R., Junker, B., & Bickel, D. D. (2010). Investigating the effectiveness of a comprehensive literacy coaching program in schools with high teacher mobility. *Elementary School Journal, 111*(1), 35–62.

Moats, L. C. (1994). The missing foundation in teacher education: Knowledge of the structure of spoken and written language. *Annals of Dyslexia, 44,* 81–102.

National Center for Education Statistics. (2019). *National Assessment of Educational Progress (NAEP), National achievement-level results: NAEP report card: Reading.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

OMG Center for Collaborative Learning. (2009). Model classroom project evaluation: Final Report. Philadelphia.

Papay, J. P., Bacher-Hicks, A., Page, L. C., & Marinell, W. H. (2017). The challenge of teacher retention in urban schools: Evidence of variation from a cross-site analysis. *Educational Researcher, 46*(8), 434–448.

Parkinson, J., Salinger, T., Meakin, J., & Smith, D. (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.* Washington, DC: American Institutes for Research.

Piasta, S. B., Connor, C. M. D., Fishman, B. J., & Morrison, F. J. (2009). Teachers' knowledge of literacy concepts, classroom practices, and student reading growth. *Journal of Poverty, 13*(3), 224–248.

Shanahan, T., & Lonigan, C. (2010). The National Early Literacy Panel: A summary of the process and the report. *Educational Researcher, 39*, 279–285.

Smith, J. L., Nelson, N. J., Fien, H., Smolkowski, K., Kosty, D., & Baker, S. K. (2016). Examining the efficacy of a multitiered intervention for at-risk readers in grade 1. *The Elementary School Journal, 116*(4), 549–573.

Smith, M. W., Brady, J. P., & Clark-Chiarelli, N. (2008). *Early language and literacy classroom observation K–3 tool* (Research ed.). Baltimore, MD: Paul H. Brookes.

Teacher Quality: Reading and Writing Assessment grant (R305A100641), funded by the National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*(1), 3–32.

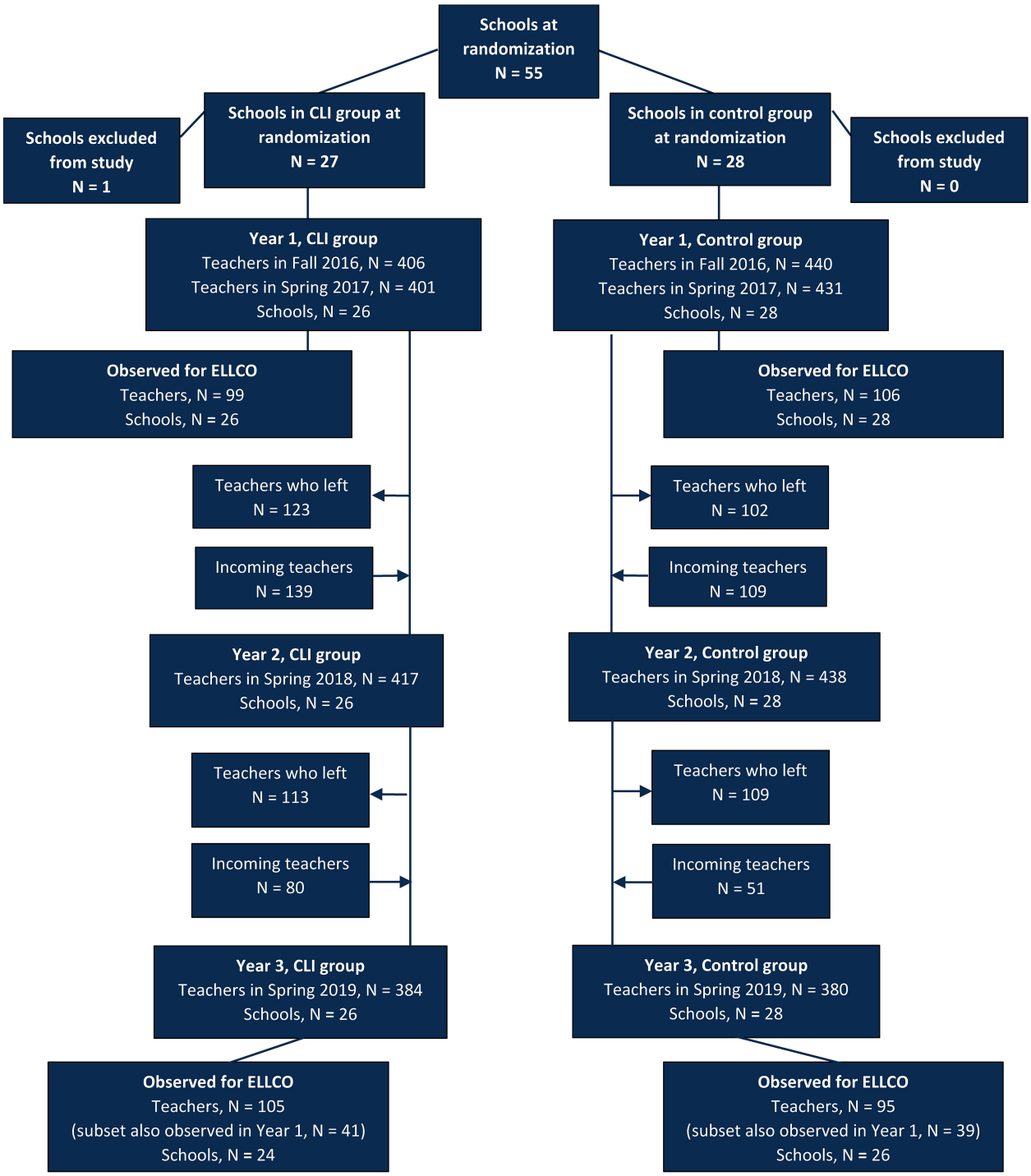
Williams, K. T., Cassidy, J., & Samuels, S. J. (2001). *GRADE teacher's scoring and interpretive manual level 1*. San Antonio, TX: Pearson Education.

Appendix A. Details About the Study Sample

This appendix presents additional details about the sample. First, we present the flow of the schools and teachers in the study. Next, we present details on the student sample and attrition for the study-administered GRADE assessment. We also provide information on the English proficiency assessments that were used as criteria for EL students within the GRADE subsample. Finally, we present completion rates for the samples of teachers who completed the study-administered survey each year.

Sample of Schools and Teachers

Figure A-1. School and Teacher Sample Flow



Sample of Students for the GRADE Assessment

Table A-1 shows size of the sample for students who were randomly selected to take the GRADE assessment. Sample size is shown over time, separately by entry grade for treatment and control conditions. Four factors affected the analytic sample of those students who had been selected in fall of Year 1 to take the GRADE assessment: (1) while most families provided consent for their student to participate in the GRADE assessment; some did not; (2) most students with positive consent took the GRADE, but some did not due to absence or other conflicts on testing days; (3) each year, schools experienced attrition of students who were then no longer in their original school; and (4) each year, some students aged out of the tested grade levels.

Table A-1. GRADE Sample Size by Entry Grade

Number of Students	Entry Grade[a]							
	Kindergarten		Grade 1		Grade 2		Grade 3	
	T	C	T	C	T	C	T	C
Sampled for GRADE assessment	1,031	1,087	979	1,077	1,052	1,103	1,068	1,119
Consented to participate in GRADE assessment	793	816	753	797	809	829	805	805
With baseline GRADE pretest scores[b]	575	567	573	566	553	545	554	522
With Year 1 GRADE posttest scores	695	694	660	718	700	730	737	725
With Year 2 GRADE posttest scores	581	580	514	555	568	610	n/a	n/a
With Year 3 GRADE posttest scores	497	513	444	493	n/a	n/a	n/a	n/a

Source: Study-administered GRADE assessment

Notes. T = treatment, C = control; n/a = not applicable, due to grade progression.

[a]*Entry grade* refers to the grade level of students when they entered the study in Year 1.

[b]Students with missing baseline scores may still contribute outcome data with their baseline score set equal to the mean non-missing score within a school and grade.

Table A-2 presents the overall and differential attrition rates for the subsample of students taking the GRADE assessment at the end of each intervention year.

Table A-2. Overall and Differential Attrition Rates for Student Achievement Outcomes Measured Based on the GRADE Assessment, By Year

Year	Overall Attrition Rate	Attrition Rate for Treatment Group	Attrition Rate for Control Group	Differential Attrition Rate[a]
Year 1	33.5%	32.4%	34.6%	-2.2%
Year 2	46.2%	45.7%	46.6%	-0.9%
Year 3	53.4%	53.1%	53.5%	-0.4%

Source: Study-administered GRADE assessment

[a]Differential rate may not match the difference between the displayed rates in Columns 3 and 4 due to rounding.

Because delivering a study-administered literacy assessment in multiple languages was beyond the scope of this study, only students who met their district's requirement for testing in English were eligible for random selection into the GRADE subsample. Students' English proficiency was based on outcomes from state-specific English proficiency assessments (see Table A-3), using the same criterion the district used in 2016 to identify students' language of assessment. The bottom row in Table A-3 shows the rate of ELs in the GRADE subsample in each study district. Other outcomes for ELs were examined through the English proficiency assessment outcomes (see Appendix B).

Table A-3. Selection of EL Students for Study-Administered GRADE Assessment in Each District

	Broward	Denver	Elizabeth	Houston
Assessment Used to Determine Eligibility for GRADE testing, Grades 1–3 (Spring 2016)	ACCESS	ACCESS	ACCESS	TELPAS
Assessment Used to Determine Eligibility for GRADE Testing, Kindergarten and Students New to the District (Fall 2016)	IPT	ACCESS Placement	ACCESS Placement	IPT
EL Rate for K–3 Students in Sample Schools (2016–17)	35.4%	56.1%	33.6%	53.9%
EL Rate in Study Subsample for GRADE Assessment (2016–17)	24.3%	35.7%	22.8%	32.8%

ACCESS = WIDA Consortium's Assessing Comprehension and Communication in English State-to-State; TELPAS = Texas English Language Proficiency Assessment System; IPT = IDEA Proficiency Test

Teacher Survey Response Rates

The survey sample included the pool of K–3 teachers each year who taught literacy in the study schools. Teacher survey response rates averaged 61% across years and conditions. Table A-4 shows the number of survey participants and response rate separately for treatment and control teachers for each survey administration.

Table A-4. Completion Rates for the Teacher Survey in Years 1 Through 3

Timing of Survey	Treatment Teachers			Control Teachers		
	Total N Surveyed	N of Respondents	Response Rate	Total N Surveyed	N of Respondents	Response Rate
Fall Year 1	404	273	68%	438	268	61%
Spring Year 1	401	198	49%	431	216	50%
Spring Year 2	429	272	63%	404	255	63%
Spring Year 3	384	272	71%	381	235	62%

Source: Study-administered survey of teachers

Appendix B. Findings From Exploratory Analyses

This appendix presents findings from a series of analyses designed to answer the exploratory research questions:

- What is the impact of the CLI intervention on particular subskills of reading?
- To what degree is there a differential impact on the basis of students' baseline achievement?
- What is the impact of the CLI intervention for students whose teachers had remained in the study schools for 3 years?
- What is the impact of the CLI intervention on EL students' English proficiency outcomes?
- What is the impact of the CLI intervention on teachers' knowledge of beginning reading?

Results are based on multilevel HLM models similar to those used for the primary analyses in the main body of the report.

Exploratory Analyses Based on the GRADE Assessment Data

For the first two exploratory questions, we conducted analyses of achievement outcomes on the GRADE to examine (a) whether the CLI intervention had an impact on specific literacy skills as measured by GRADE subtests and (b) whether the intervention had an impact on students in subgroups based on baseline reading achievement. These issues were explored given findings from the prior validation study of the CLI intervention. The previous study suggested that, among one cohort, the intervention was more effective for students with above-average baseline achievement in Year 1, with those differences fading by Year 2. For another cohort, there was significant differential effect favoring low-achieving students. The prior study also showed that the positive results of the intervention for younger students (kindergarteners) were largely driven by statistically significant impact on students' letter-word reading skills. For Grade 2 students, the impact of the intervention was larger and statistically significant for comprehension skills compared to word reading and meaning.

Table B-1 presents estimates of the effects of the CLI intervention on each of the reading subtests that compose the GRADE test. Only one subtest effect is statistically significant: the effect on the vocabulary subtest given to students in Grade 3 (ES = 0.17, $p < .05$). Considering that the point estimates for vocabulary in both Year 1 and Year 2 are negative, the Year 3 result does not imply that CLI is particularly strong at promoting vocabulary skills. The students enrolled in Grade 3 in Year 3 had relatively larger overall treatment effects across years, as shown in Tables 9 and 10, so the vocabulary effect most likely reflects subgroup, grade-level effects.

Table B-1. Effects of the CLI Intervention on GRADE Subtest Scores, by Year

GRADE Subtest	Statistics	Year 1	Year 2	Year 3
Word Meaning (Subtest 1, Grades 1–2)	ES	-0.007	0.013	-0.013
	p-value	0.897	0.841	0.855
	Sample size	2,873	2,238	1,016
	Treatment	1,398	1,099	499
	Control	1,475	1,139	517
Vocabulary (Subtest 1, Grade 3)	ES	-0.075	-0.046	0.174*
	p-value	0.300	0.516	0.013
	Sample size	1,486	1,179	938
	Treatment	749	568	445
	Control	737	611	453
Word Reading (Subtest 2, Grades 1–3)	ES	0.005	-0.012	-0.001
	p-value	0.878	0.712	0.989
	Sample size	4,359	3,417	1,954
	Treatment	2,147	1,667	944
	Control	2,212	1,750	1,010
Sentence Comprehension (Subtest 3, Grades 1–3)	ES	0.009	-0.010	0.033
	p-value	0.831	0.848	0.608
	Sample size	4,359	3,417	1,954
	Treatment	2,147	1,667	944
	Control	2,212	1,750	1,010
Passage Comprehension (Subtest 4, Grades 1–3)	ES	0.03	0.011	0.007
	p-value	0.949	0.829	0.930
	Sample size	4,359	3,417	1,954
	Treatment	2,147	1,667	944
	Control	2,212	1,750	1,010

Source: Study-administered GRADE assessment

Note. *P*-values are based on two-tailed *t* tests; *p<.05; **p<.01

Table B-2 presents the impact results for students in different baseline achievement terciles. There were statistically significant positive effects of the CLI intervention for students in the highest tercile in Year 2. However, the pattern of results by baseline achievement is not consistent across years. Furthermore, none of the estimated effects per tercile are statistically different from estimated effects for other terciles in the same year.

Table B-2. Effects of the CLI Intervention on Overall GRADE Scores, by Year and Tercile of Baseline Achievement

Baseline Achievement Tercile	Statistics	Year 1	Year 2	Year 3
Lowest Tercile	ES	0.011	-0.007	-0.050
	p-value	0.902	0.737	0.676
	Sample size	1,357	877	590
	Treatment	689	434	302
	Control	668	443	288
Second Tercile	ES	0.027	0.003	0.100
	p-value	0.742	0.963	0.425
	Sample size	1,349	745	414
	Treatment	696	377	206
	Control	653	368	208
Highest Tercile	ES	0.076	0.206**	0.128
	p-value	0.188	0.002	0.219
	Sample size	1,292	806	369
	Treatment	641	400	182
	Control	651	406	187

Source: Study-administered GRADE assessment

Notes. ESs were calculated using the pooled within-group standard deviation. *P*-values are based on two-tailed *t* tests. *p<.05; **p<.01

Next, for the third exploratory question, we analyzed impact findings based on GRADE achievement data for a subset of students whose teachers had been in the same schools since the beginning of the study. Given the relatively high amount of teacher attrition, we wanted to explore the effect of the CLI intervention for treatment students with teachers who had the opportunity to receive greater exposure to the CLI intervention. Therefore, we restricted the achievement impact sample to students of stable teachers in both conditions. Specifically, students in the Year 2 analysis sample included 900 Grade 1 students, 812 Grade 2 students, and 891 Grade 3 students who had "stable" Year 2 teachers who had also been in the same study schools in Year 1. Students in the Year 3 analysis sample included 483 Grade 2 students and 432 Grade 3 students who had Year 3 teachers who had also been in the same schools in Years 1 and 2. As Table B-3 shows, there were positive and statistically significant effects in Year 2 for students in Grade 2 (ES = .19, $p < .01$) and in Year 3 for students in Grade 3 (ES = .35, $p < .01$). Both of these estimates involve the students who entered the study in Grade 1. In results previously presented in the main report, this same age group had larger, though not statistically

significant, point estimates compared to other grade levels in the analyses that were not limited to only students with stable teachers (see Table 9). The results here suggest that students with stable teachers (and treatment teachers with greater exposure to the CLI intervention) are contributing to that result.

Table B-3. Effects of the CLI Intervention on the Overall GRADE Scores for Students of Stable Teachers, by Year and Grade

Grade Level at Time of Assessment	Statistics	Year 2	Year 3
Grade 1	ES	-0.044	n/a
	p-value	0.647	
	Student sample	900	
	Treatment (Schools)	416 (25)	
	Control (Schools)	484 (28)	
Grade 2	ES	0.192**	0.174
	p-value	0.004	0.224
	Student sample	812	483
	Treatment (Schools)	391 (26)	210 (21)
	Control (Schools)	421 (28)	273 (24)
Grade 3	ES	0.008	0.351**
	p-value	0.943	0.007
	Student sample	891	432
	Treatment (Schools)	423 (25)	203 (19)
	Control (Schools)	468 (27)	229 (20)

Source: Study-administered GRADE assessment

Notes. ESs were calculated using the pooled within-group standard deviation. *P*-values are based on two-tailed *t* tests. *p<.05; **p<.01

Exploratory Analyses Using Extant English Proficiency Data for the EL Subsample

Because the sample contained a relatively high percentage of ELs, it was of interest to CLI to consider reading outcomes particular to the EL subgroup. To explore this, we also conducted exploratory analyses to examine the impact of the CLI intervention on EL students' English language proficiency in each intervention year. We collected extant state English language proficiency assessment data. Data included reading domain outcomes from the Assessing Comprehension and Communication in English State-to-State for English Language Learners (ACCESS) for Broward, Denver, and Elizabeth as well as the Texas English Language Proficiency

Assessment System (TELPAS)[5] for Houston. The K–3 students who were identified as ELs by the districts in fall of Year 1 were included in the EL outcome analysis. We standardized the English proficiency test scores based on the state means and standard deviations within year and grade to place data from different states on a common metric. The difference between treatment and control groups on English proficiency reading domain scores was not statistically significant in any of the 3 intervention years, as shown in Table B-4.

Table B-4. Effects of the CLI Intervention on the English Language Proficiency of ELs, by Year

Statistics	Year 1	Year 2	Year 3
ES	-0.079	0.016	0.062
p-value	0.146	0.807	0.220
Sample size	5,770	5,577	4,915
Treatment	2,540	2,438	2,157
Control	3,230	3,139	2,758

Source: District-provided English language proficiency data
Note. *P*-values are based on two-tailed *t* tests.
*p<.05; **p<.01

Exploratory Analyses of Impact on Teacher Knowledge

For the final exploratory question, we examined the impact of the CLI intervention on teacher knowledge of beginning reading as measured by the Teacher Knowledge of Student Content Engagement (TK-SCE) assessment (Teacher Quality: Reading and Writing Assessment grant, R305A100641). This assessment was designed and pilot-tested for validation under a federally funded measurement grant.[6] The measure outlines three factors involved in the teaching and learning processes as young children learn to read: (*a) knowledge of content* or broad concepts teachers should know about reading as well as the prerequisite linguistic knowledge students need; (b) *knowledge of pedagogy* specific to literacy instructional activities, strategies, and assessment to foster learning; and (c) *knowledge of students* or what teachers should know about student readiness, prior knowledge, motivation, and engagement with reading and

[5] The TELPAS Reading domain is not administered to kindergarten and Grade 1 students; therefore, students in those grades from Houston were not included in the analysis.
[6] The TK-SCE was developed under a U.S. Department of Education, Institute of Education Sciences grant titled "Validation of an Assessment of Teacher Knowledge of Beginning Reading Instruction." For validation, the survey items were administered to teachers in 20 schools in nine districts. Their K–3 students (n = 1,399) took the Dynamic Indicators of Beginning Literacy Skills (DIBELS) test. The percentage of students at or above the benchmark cut-point on each DIBELS subtest was compared between teachers scoring above the mean and teachers scoring below the mean on each TK-SCE domain score or total score. Although no comparison yielded a statistically significant result, the percentage of students at or above the benchmark cut-points was higher on all DIBELS measures for teachers who scored above the mean across all three TK-SCE domains, on average, than for teachers who scored below the mean.

writing. In the teacher survey, the study team selected 30–35 multiple-choice items from the TK-SCE item library to represent the three factors covered by the TK-SCE.

Treatment and control teachers were equivalent in teacher knowledge scores at baseline. There were no statistically significant differences between the groups at the end of Years 1, 2, or 3 (see Table B-5).

Table B-5. Effects of CLI Intervention on Teachers' Knowledge, by Year

Teacher Knowledge Area	Statistics	Year 1	Year 2	Year 3
Content	ES	-0.002	0.007	-0.185
	p-value	0.983	0.935	0.037
Pedagogy	ES	0.196	-0.046	-0.169
	p-value	0.049	0.592	0.064
Students	ES	0.070	-0.090	-0.077
	p-value	0.463	0.295	0.386
Teacher sample size	Treatment	198	272	271
	Control	216	255	236

Source: Study-administered teacher survey

Note. Sample contained 54 schools; 26 in the CLI group and 28 in the control group in each year.

Appendix C. Implementation Fidelity Matrix

	Indicator	Operational Definition	Data Source	Level	Indicator Scoring	School-Level Score
Key Component 1 = Resources, Professional Development, and Coaching for Teachers						
1.	**Books and materials** provided to each CLI classroom	Each classroom receives designated number of books and other supplies (Year 1 only).	CLI records of classroom resources	Classroom	**0 = No,** classroom did not receive adequate materials. **1 = Yes,** classroom received adequate materials.	**1 = Low:** 0% to 50% of classrooms had "Yes." **2 = Mod:** 51% to 75 % of classrooms had "Yes." **3 = High:** 76% to 100% of classrooms had "Yes."
2.	**Early literacy instructional seminars** provided in each district	Teachers attend the two-day summer/fall seminar and the one-day seminar during school year in Year 1 and two 1-day seminars each in Years 2 and 3.	CLI records: attendance rosters	Teacher	**1 =** One day attended. **2 =** Two days attended. **3 =** All 3 days attended.[7]	**1 = Low:** School average below 2.0 **2 = Mod:** School average between 2.0 and 2.5 **3 = High:** School average above 2.5
3.	**Content coaching** provided to CLI teachers	Teachers receive designated hours of coaching: • 35 hours in Year 1 • 40 hours in Year 2 • 30 hours in Year 3	Coaching logs	Teacher	**1 =** Coaching hours are 0% to 50% of expected. **2 =** Coaching hours are 51% to 75% of expected. **3 =** Coaching hours are 76% to 100% of expected.	**1 = Low:** School average below 2.0 **2 = Mod:** School average between 2.0 and 2.5 **3 = High:** School average above 2.5
4.	**Content coaching focused** on three of the areas highlighted in the TELP checklist	Coaches customize their interactions with the teachers but cover at least three TELP areas during the course of the year (Years 1–3).	Coaching logs	Teacher	**1 =** Coaching focuses on only one area. **2 =** Coaching focuses on only two areas. **3 =** Coaching focuses on three or more areas.	**1 = Low:** School average below 2.0 **2 = Mod:** School average between 2.0 and 2.5 **3 = High:** School average above 2.5
5.	**Additional content coaching** provided to CLI teachers	Schools receive 10 additional hours of coaching (Years 1–3).	Coaching logs	School	Same as school-level scoring	**1 =** Coaching hours are 0% to 50% of expected. **2 =** Coaching hours are 51% to 75% of expected. **3 =** Coaching hours are 76% to 100% of expected.
6.	**CLI-facilitated grade-level meetings** for teachers	Teachers attend eight CLI-facilitated meetings per school year, at which model lessons are crafted, implemented, and revised (Years 1–3).	CLI records: attendance rosters and meeting notes	Teacher	**1 =** Teacher attends fewer than five meetings. **2 =** Teacher attends five or six meetings. **3 =** Teacher attends seven or eight meetings.	**1 = Low:** School average below 2.0 **2 = Mod:** School average between 2.0 and 2.5 **3 = High:** School average above 2.5

School-level score for Component 1 is the average of Indicators 1–6.

Sample-level score for Component 1 is the average of school-level scores: ***1 = Low:*** *average below 2.0;* ***2 = Mod:*** *average between 2.0 and 2.5;* ***3 = High:*** *average above 2.5*

Sample-Level Fidelity Met = Average above 2.5

[7] In Years 2 and 3, the indicator scoring is adjusted as follows: 1 = 0 days attended, 2 = 1 day attended, 3 = both days attended.

	Indicator	Operational Definition	Data Source	Level	Indicator Scoring	School-Level Score
Key Component 2 = Instructional Lead Professional Development and Coaching						
7.	**Selection of an instructional lead teacher (ILT) from pool of K–3 teachers**	Two teachers per practice in each school are identified as an ILT. Each school selects the expected number of ILTs each year: • 4 ILTs in Year 1 • 6 ILTs in Year 2 • 4 ILTs in Year 3	CLI data: List of ILTs by school	Teacher	**0 = No**, ILT not appointed. **1 = Yes**, ILT is appointed.	**1 = Low:** 50% or fewer of expected ILT teachers. **2 = Mod:** Above 50% and below 100% of expected ILT teachers. **3 = High:** 100% of expected ILT teachers.
8.	**Additional coaching** and support provided to ILTs	ILTs receive additional coaching: • 5 hours in Year 1 • 10 hours in Year 2 • 10 hours in Year 3	Coaching logs	Teacher	**1 =** Additional coaching hours are 0% to 50% of expected. **2 =** Additional coaching hours are 51% to 75% of expected. **3 =** Additional coaching hours are 76% to 100% of expected.	**1 = Low:** School average below 2.0 **2 = Mod:** School average between 2.0 and 2.5 **3 = High:** School average above 2.5
9.	**ILT seminars**	ILTs will attend 2 seminars that discuss the ILT's assigned area of focus (Years 1–3).	CLI records: attendance rosters	Teacher	**1 =** Zero seminars attended with ILT's assigned model focus. **2 =** One seminar attended with ILT's assigned model focus. **3 =** Two seminars attended with ILT's assigned model focus.	**1 = Low:** School average below 2.0 **2 = Mod:** School average between 2.0 and 2.5 **3 = High:** School average above 2.5

School-level score for Component 2 is the average of Indicators 7–9.

Sample-level score for Component 2 is the average of all school-level scores: ***1 = Low:*** *average below 2.0;* ***2 = Mod:*** *average between 2.0 and 2.5;* ***3 = High:*** *average above 2.5*

Sample-Level Fidelity Met = Average above 2.5

	Indicator	Operational Definition	Data Source	Level	Indicator Scoring	School-Level Score
Key Component 3 = Administrative Professional Development						
10.	**Leadership Team Meetings**	Leadership team meetings held four times a year. Attendees include one or more school-based administrators, and the CLI regional manager (Years 1–3).	CLI records: attendance rosters	School	Same as school-level scoring	**1 = Low:** Administrator attendance at two or fewer meetings **2 = Mod:** Administrator attendance at only three meetings **3 = High:** Administrator attendance at all four meetings
11.	**Principal meetings**	Three principal meetings are held during the school year. Attendees include principals of all CLI schools in each district and a CLI representative (Years 1–3).	CLI records: attendance rosters	District	**1 = Low:** Average attendance among principals below 50% across all three meetings (or any meeting not held) **2 = Mod:** Average attendance among principals between 50% and 75% across all three meetings **3 = High:** Average attendance among principals above 75% across all three meetings	n/a
12.	**Review-of-Progress** meetings	Two review-of-progress meetings, one held at the beginning and one held at the end of the year. Attendees include at least one representative from each school, at least one district-level representative, and a CLI representative (Years 1–3).	CLI records: attendance rosters	District	**1 = Low:** Attendance below 50% **2 = Mod:** Attendance between 50% and 75% **3 = High:** Attendance above 75%	n/a

School-level score for Component 3 is the average of Indicators 10–12.

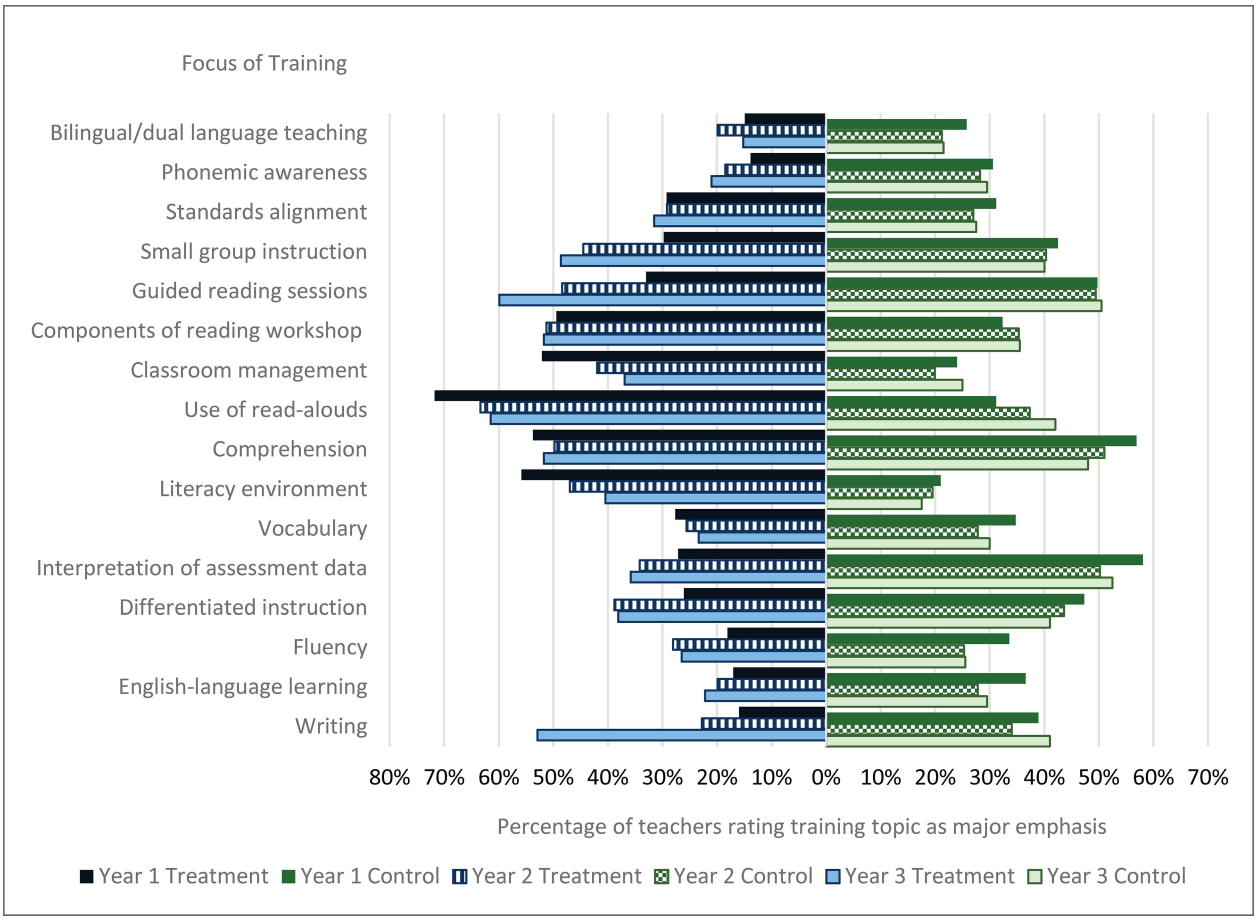
*Sample-level score for Component 3 is the average of all school-level scores: **1 = Low**: average below 2.0; **2 = Mod**: average between 2.0 and 2.5; **3 = High**: average above 2.5*

Sample-Level Fidelity Met = Average above 2.5

Appendix D. Teacher Ratings of Professional Development Content

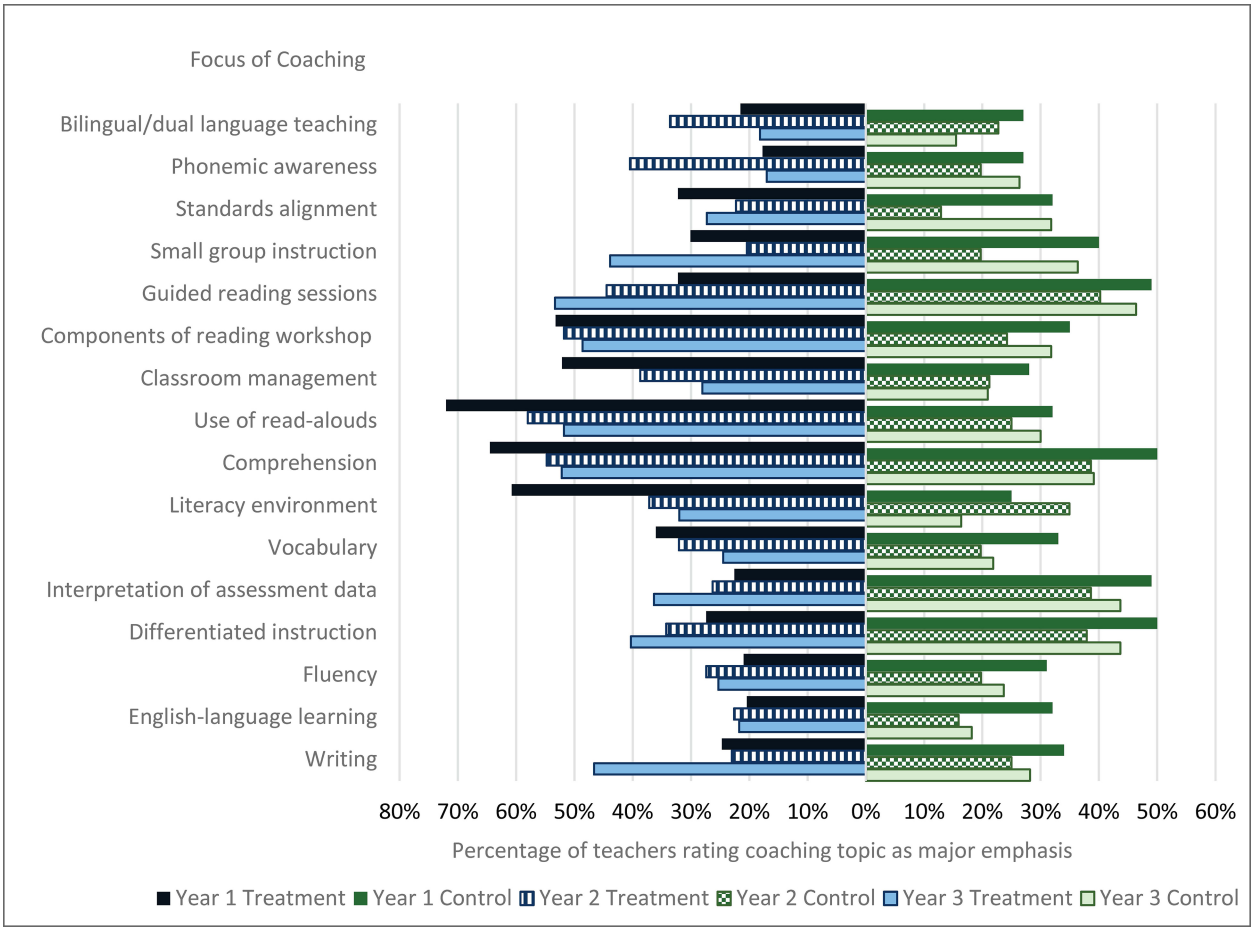
This appendix provides additional information from the teacher survey about teacher ratings of PD content. Teachers responded to survey questions that asked them to rate the degree to which various literacy topic areas were emphasized in the PD they received each year. The two figures below show the percentage of teachers in each study group who rated each topic area as a "major emphasis" of PD. Figure D-1 shows teachers' ratings of different types of training activities (e.g., coursework, institutes, workshops, learning communities) in which they participated in each of the 3 intervention years. Figure D-2 shows teachers' ratings of different types of coaching activities in which they participated in each year.

Figure D-1. Percentage of Teachers Rating Each Training Activity as a "Major Emphasis" of the Professional Development They Received, by Year and Study Condition



Source: Study-administered survey of teachers

Figure D-2. Percentage of Teachers Rating Each Coaching Activity as a "Major Emphasis" of the Professional Development They Received, by Year and Study Condition



Source: Study-administered survey of teachers

Appendix E. Baseline Equivalence for Student Achievement Analysis Samples

This appendix presents baseline equivalence information for the samples used in the analyses of the GRADE and state ELA assessment data.

Baseline Equivalence for the GRADE Analysis Samples

Table E-1 presents the baseline differences between the two study conditions for all students selected to take the GRADE as well as for students in the analytic samples for the GRADE assessment. Differences in baseline GRADE scores are expressed in terms of standard deviations; all other variables assessed at baseline are binary, and differences in those variables are expressed in terms of Cox's index. Each is based on a regression of the given baseline variable on block fixed effects and treatment-district interactions.

In all 3 years, there were no significant differences between conditions in terms of baseline GRADE test scores (see Table E-1). Treatment students in the GRADE analytic samples were less likely to have missing baseline GRADE scores than control students. Table E-1 also illustrates that, compared with control students, treatment students in the GRADE analytic sample were less likely to be eligible for subsidized lunch at baseline and less likely to have been classified as ELs or as having a disability, though the differences between the two groups were not always statistically significant in all 3 years. There were not significant differences between conditions in terms of the likelihood of students belonging to a racial or ethnic minority group.

Table E-1. Baseline Equivalence for Students in the Analytic Sample for GRADE in Each Year

Measure	Baseline Difference, Year 1 Sample	Baseline Difference, Year 2 Sample	Baseline Difference, Year 3 Sample
Baseline GRADE score	-0.033	-0.016	-0.079
Baseline GRADE score missing	-0.175**	-0.144**	-0.237**
Subsidized lunch indicator	-0.177**	-0.245**	-0.152**
EL indicator	-0.053	-0.126**	-0.094
Disability indicator	-0.299**	-0.205*	-0.245
Minority indicator	0.065	0.045	0.016
Sample size	5,659	3,409	1,949
Treatment	2,792	1,664	943
Control	2,867	1,745	1,006

Source: Study-administered GRADE assessment

Note. *P*-values are based on two-tailed *t* tests.

*p<.05; **p<.01

Baseline Equivalence for the State ELA Assessment Analysis Samples

Table E-2 presents the baseline differences in the analytic samples for the standardized state ELA assessment scores. Each year additional grade levels of students were added to the sample as they went into Grade 3 and took the ELA assessments. The Year 1 sample consists of only students who were in Grade 3 in Year 1; the Year 2 sample adds another grade level of students who were in Grade 2 in Year 1, and the Year 3 sample adds one more grade level of students who were in Grade 1 in Year 1. The baseline differences resemble those for the GRADE subsample; there were no significant differences between conditions in terms of baseline GRADE scores. Students in treatment schools were less likely to have missing baseline scores, less likely to be eligible for subsidized lunch, and in Year 2, were less likely to have a disability. There were no significant differences in terms of whether students belong to a minority racial or ethnic group.

Table E-2. Baseline Equivalence for Students in the Analytic Samples for State ELA Assessment

Measure	Baseline Difference, Year 1 Sample	Baseline Difference, Year 2 Sample	Baseline Difference, Year 3 Sample
Baseline GRADE score	0.032	0.086	-0.014
Baseline GRADE score missing	-0.107*	-0.146**	-0.150**
Subsidized lunch indicator	-0.111*	-0.147**	-0.138**
EL indicator	-0.064	-0.118**	-0.080**
Disability indicator	-0.079	-0.159**	-0.087
Minority indicator	-0.050	-0.068	-0.027
Sample size	4,122	6,542	8,622
Treatment	1,978	2,996	3,980
Control	2,144	3,546	4,642

Source: Study-administered GRADE assessment

Note. *P*-values are based on two-tailed *t* tests.

*p<.05; **p<.01

Acknowledgements

The AIR study team has appreciated our collaboration with the Children's Literacy Initiative on the i3 evaluation. First, we would like to thank the districts—Broward County Public Schools in Florida, Denver Public Schools in Colorado, Elizabeth Public Schools in New Jersey, and the Houston Independent School District in Texas—and the student study participants, parents who provided consent, teacher study participants and the district and data office staff who helped coordinate the complexities of data collection.

Next, we also express our sincere gratitude to the staff from our study partner, School Readiness Consulting (SRC), who conducted data collection in the schools. The team included Lindsey Allard Agnamba, Laura Hawkinson, BreAnna Davis Tribble, Janell Chery, Grace Wagner, and Rebecca Weiss, along with SRC's teams of district coordinators and data collectors.

Finally, we would like to thank our AIR colleagues, namely Mengli Song and Jordan Rickles, who conducted reviews of the data analysis and reports, Dong Hoon Lee, for his work on the data analysis, and Julia Parkinson, who was responsible for early design plans. We also appreciate the assistance of AIR's Publications and Creative Services, especially editor Christine Hulbert.

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Established in 1946, with headquarters in Arlington, VA, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance, both domestically and internationally, in the areas of education, health, and the workforce. For more information, visit www.air.org.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3239 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Arlington, VA (HQ) | Sacramento and San Mateo, CA | Atlanta, GA | Chicago, IL | Indianapolis, IN | Waltham, MA Frederick and Rockville, MD | Chapel Hill, NC | Cayce, SC | Austin, TX

International: Ethiopia | Haiti

12891_09/20