Assessing the Value-Added Effects of Literacy Collaborative Professional Development on Student Learning

Gina Biancarosa, Stanford University
Anthony Bryk, Stanford University
Emily Dexter, Lesley University

Paper presented at the 2008 Conference of the American Educational Research Association, New York, New York.

The work described here has been supported by a Teacher Quality Grant from the Institute for Educational Sciences, R305M040086, "Can Literacy Professional Development be improved with Web-Based Collaborative Learning tools: A Randomized Field Trial." We are appreciative of the support provided by IES. All errors of fact, omission and/or interpretation are solely the responsibility of the authors.

The use of school-based literacy coaches as a professional development strategy has become widespread in U.S. schools. For example, many large urban districts have made large investments in initiatives to train and support literacy coaches. While an extensive literature advocates for this approach, few empirical studies of coaching and its effects on teaching practice and student achievement exist. This paper examines the value-added effects of literacy coaching on student literacy learning based on the first three years of a four-year longitudinal field trial of the effectiveness of Literacy Collaborative (LC), which relies heavily on literacy coaching as a means of improving student literacy learning.

Background

There is a growing body of educational literature on coaching and on literacy coaching in particular. Most of it is descriptive and prescriptive and not grounded in formal quantitative or qualitative research. Walpole and McKenna (2004), Toll (2007), Blachowicz et al. (2005), Allen (2006), Casey (2006), and Bean and Carol (2006), for example, offer descriptions of the role of the literacy coach and how best to fulfill that role. The International Reading Association (2004, 2006) describes qualifications and ability standards for literacy coaches. However, as Neufeld and Roper (2003) note, “No one, as yet, has proven that coaching contributes significantly to increased student achievement. Indeed, there are scant studies of this form of professional development and how it influences teachers’ practice and students’ learning" (p.1).

Research on Literacy Coaching

Most of the empirical literature on literacy coaching is in the form of program evaluations rather than basic research studies, and most is grounded in qualitative methods (Gibson, 2006; Neufeld and Roper, 2003). Neufeld and Roper (2003), for example, used qualitative data to describe the actual work of coaches in four urban districts. Poglinco, et al. (2003) conducted a descriptive study of coaching in 27 schools that implemented a comprehensive school reform model. Although they used a four-point rubric to evaluate teachers' practices against program standards, they did not measure change in teacher practice or assess effects of coaching on these practices, and most critically they did not examine effects of either coaching or teacher practices on student learning.

The most comprehensive program evaluation of coaching conducted to-date is a three-year study by the Rand Corporation (Marsh et al., 2005). Two of three urban school districts they studied placed full-time English Language Arts coaches in all their schools. Marsh et al. examined changes in the schools' ELA proficiency percentages over multiple years. One of the districts with coaches showed "substantial" improvement in district scores, while the other showed "limited" improvement. However, even though teachers reported benefits from coaching, Marsh et al. did not examine student-level scores and did not test causal models of the effects of coaching.

Several other evaluations of literacy coaching programs have been conducted, including in Alabama (Norton, 2007), Alaska (Barton & Lavrakas, 2006), and Idaho (Reed & Rettig, 2005). None, however, have offered rigorous empirical evidence of the effects of coaching on teacher change or student literacy growth.

Literacy Collaborative

Established in 1993, Literacy Collaborative is a comprehensive school reform program designed to improve elementary children's reading, writing, and language skills in large part through school-based coaching. The program builds on 30 years of research and development grounded in the reading theories of Marie Clay (1979, 2004) and elaborated by Fountas and

Pinnell (1996, 2006). A key component of Literacy Collaborative is the training and support of school-based literacy coaches, who are teachers selected by their schools to lead local instructional improvement efforts. Coaches receive rigorous training in the theory and content of literacy learning, as well as in how to teach children within Literacy Collaborative's instructional framework. Once trained, the coaches, known as Literacy Coordinators, offer extensive school-based professional development activities, including individual coaching, in order to improve teaching practices within a framework of comprehensive literacy instruction. The overall goal is to improve the reading and writing achievement of all children in the school.

Research Design

This paper reports preliminary findings from the first two years of a longitudinal study that seeks to investigate the impact of Literacy Collaborative professional development on student literacy learning. (The collection of data on third year effects has just been completed and these analyses will be reported in a subsequent paper.) The study is based on an accelerated longitudinal cohort design with a value-added analysis of program effects at both the school and teacher levels. During the first year of the study, literacy coaches were in training for their new role and no literacy professional development activities were conducted by them at their respective schools. Thus, the first year of this four-year study represents a "no treatment" period and affords data on student achievement for each school and classroom prior to program initiation. Implementation of LC professional development in kindergarten through second grade classrooms began in year 2, and the analyses reported in this paper focus on the first and second year effects on student literacy learning in these grades. The study continues to follow all sample students through grade 3 in order to evaluate sustaining effects of the LC program through this critical grade, where students' achievement scores typically appear for the first time in state and local district assessment-accountability systems. These follow-up results will be reported separately in a future paper.

An Accelerated Longitudinal Cohort Design

The research design is based on collecting student achievement data at four grades (K-3) over the course of four years. As a result, the study involves children from seven different cohorts who entered at different grades and in different years. Figure 1 depicts these cohorts and the timing of LC implementation for each. For example, Cohort Four entered the study in kindergarten during the first year of the study, and attended first, second, and third grade in the second, third, and fourth years of the study. In contrast, Cohort Five began in kindergarten during the second study year and attended first and second grade in the third and fourth years respectively.

Data from the first three waves of data collection (fall and spring of year one and fall of year two) offer baseline information about student growth because they are prior to LC implementation. This is denoted with solid black lines in Figure 1. Student achievement at subsequent time points, during which LC was implemented, are compared against this baseline data. The first year of LC effects are denoted by dashed lines in Figure 1, the second year by alternating dashed-and-dotted lines, and the third year by dotted lines. The fact that there is no program implementation in third grade is reflected in the figure by the third grade lines retaining the same notation as the previous year's second grade class. The effects observed in this grade are considered follow-up effects and therefore are depicted in gray.

Figure 1. Accelerated multiple cohort design: Seven cohorts in four grades across four years implementing Literacy Collaborative in grades K-2.

	Year of Study							
	Year 1 (baseline year)		Year 2 (first year of implementation)		Year 3 (second year of implementation)		Year 4 (current and final year of implementation)	
Grade	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
K	C4	C4	C5	C5	C6	C6	C7	C7
1	C3	C3	C4	C4	C5	C5	C6	C6
2	C2	C2	C3	C3	C4	C4	C5	C5
3	C1	C1	C2	C2	C3	C3	C4	C4

Baseline data
First year implementation data
Second year implementation data
Third year implementation data *(not reported here)*
Third grade follow-up effects *(not reported here)*
C# Cohort number

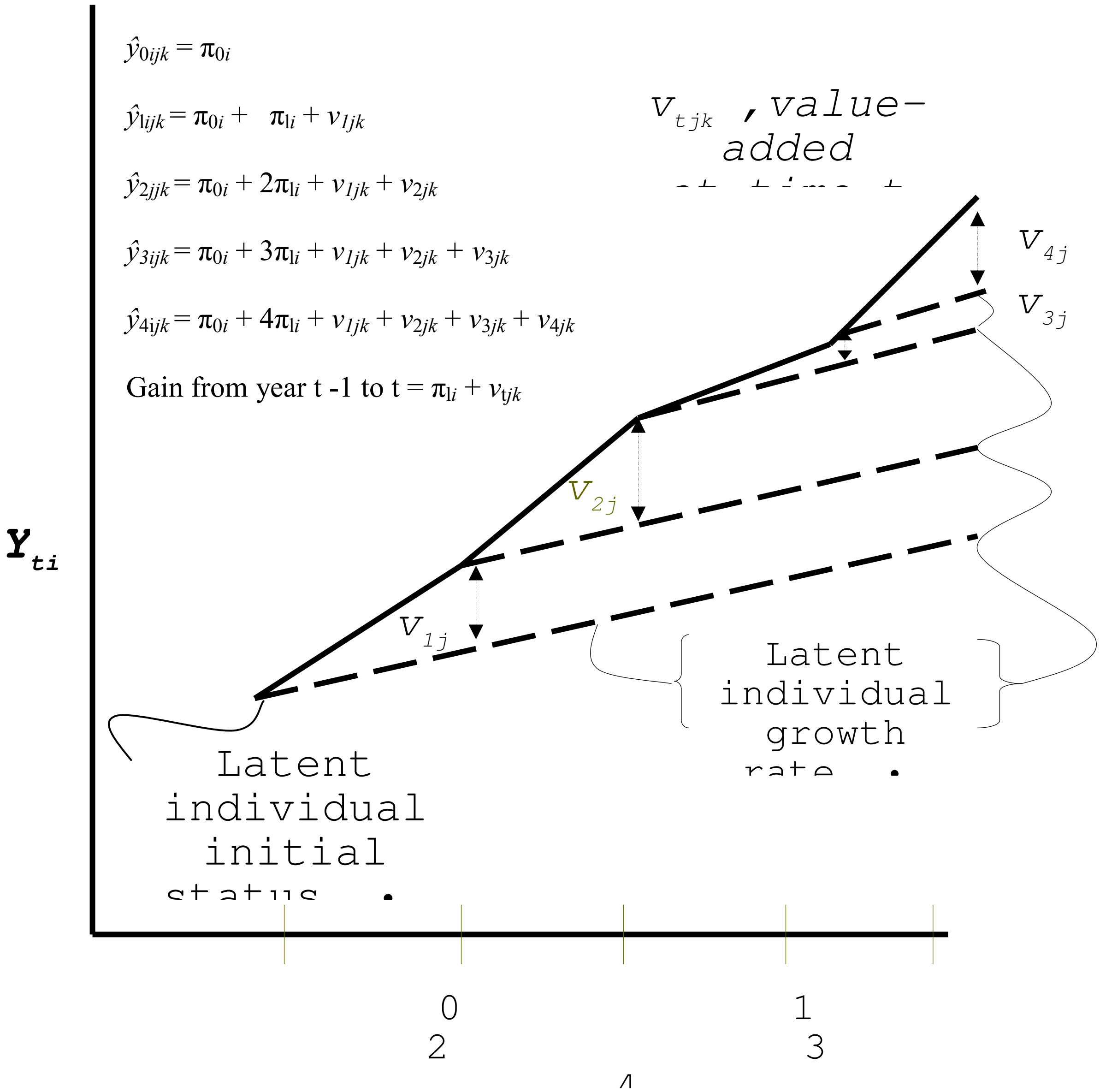
Logic of Value-Added Modeling

Our application of value-added modeling is rooted in the idea that each child has a latent growth trajectory. For simplicity, we assume that these individual trajectories can be represented as a linear function of time. The parameters in these trajectories, π_i, describe the expected achievement growth for child *i* absent the intervention. Specifically, it represents the achievement growth expected in grades K-2 for a child if he or she were exposed to the average instructional conditions prevalent in each respective school during the baseline period.

Once an intervention begins, we estimate how the actual observed student growth trajectories differ from the expected (or latent) growth trajectories under baseline conditions.

That is, we compare the observed outcome for each student *i* given exposure to a particular teacher *j* in school *k* in year *t* to the expected outcome given that student's latent growth trajectory. The value-added, which we denote as v_{tjk} in Figure 2, is the difference between these observed and expected outcomes. In general, each teacher and school may have a unique value-added effect in each time period. We are particularly interested in the value-added effects in years two and three of our study as they include potential effects associated with years one and two of LC program implementation.

Figure 2. A heuristic illustration of value-added estimation.



Participants

The study included students and teachers from18 schools in eight states across the eastern U.S.

Students. Approximately 1300 students have been assessed in each grade in kindergarten through third grade in each year of the study. This represents a student participation rate of 90%

or higher at each testing occasion. Across the schools, 40% of participating students are low income, 16% are African American, 7% are Latino, 4% are Asian American, less than 1% is Native American. Approximately 4% of students were designated as limited English proficient. See Table 1 for school-by-school demographics from the baseline year of the study.

Overall, approximately 60 percent of the student sample has complete data. These students have test scores at every occasion for which they were eligible to be assessed given their cohort (see Table 2). Of the students with incomplete data, most either entered a study school after we began data collection for their cohort or transferred out of a study school prior to third grade. In addition, some 3% missed testing on one or more occasions for which they were eligible to be assessed (i.e., due to absences).

Table 1. *Percentages of Key Demographics of the Student Sample in the Base Year.*

		School																	
	Overall	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
Low Income	40.1	33	47	20	35	36	60	19	49	38	60	50	60	73	43	45	50	86	38
Race/Ethnicity																			
African-Am.	15.5	31	3	30	1	7	41	20	1	1	41	3	1	50	0	35	0	36	4
Latino	6.9	8	31	12	2	20	2	4	0	3	2	1	0	6	0	4	0	4	2
Other	4.8	10	5	11	2	1	2	15	0	1	28	2	0	29	12	0	3	0	3
White	72.8	51	61	47	95	72	55	54	99	95	21	94	99	15	88	61	97	60	91

Note: Racial and ethnic group percentages may not total 100 due to rounding.

Table 2. *Observations per Child by Cohort for Analytic Sample (n=7971).*

Cohort	Observations							Percent with incomplete data
	1	2	3	4	5	6	Total	
2 (max = 3)	70	227	1016	0	0	0	1313	22.6
3 (max = 5)	130	263	199	191	827	0	1610	48.6
4 (max = 6)	256	367	140	245	111	783	1902	58.8
5 (max = 4)	232	373	133	992	0	0	1730	42.7
6 (max = 2)	197	1219	0	0	0	0	1416	13.9
Total	885	2449	1488	1428	938	784	7971	39.3

Teachers. A total of 366 teachers taught in kindergarten through third grade classrooms in the 18 study schools at some point during the first three years of the study. Some 207 teachers were present for all three years (or 56.6 percent), 67 (or 18.3 percent) were present for two years, and 92 (or 25.5 percent) were present for only one year. Within this group, approximately 250 kindergarten through second grade teachers participated in some form of LC professional development during this period.

Schools. As Table 1 makes clear, schools varied widely in their student composition. In several schools more than 90 percent of students were White, while in other schools 30 percent or more students were from minority groups. Similarly, the schools ranged in their socioeconomic make-up, with students receiving free or reduced price lunches ranging from a low of 19 percent to a high of 86 percent.

Schools also varied widely in the size and stability of their teaching staff over time (see Table 3). Of the 18 schools, only three (schools 2, 3, and 10) had the same number of K-3 classrooms for all three years. Even here, the three year teacher stability rates varied considerably ranging from 60-80 percent. Not surprisingly, the percentage of teachers present for all three years tended to vary even more widely among schools with a changing number of classrooms, ranging from a high of 80 percent to a low of 33 percent.

Sample adjustments. In the analyses below we focus on student results in kindergarten through second grade. The total K-2 sample includes 8,122 children and 24,086 observations. A small number of students were excluded who had repeated a grade or were tested off-grade level. This resulted in the loss of 161 children and reduced the analytic sample to 7,971 children and 23,298 observations. In addition, we excluded a small number of individual test scores with unusually large standard errors of measurement which raised questions about the reliability of these individual test administrations. This did not affect the number of children included in the analytic sample, but did reduce the number of scores by less than 2% to 22,901. The sample counts presented in Table 2 and means reported in Table 5 are based on this final analytic sample.

Table 3. *Number of Kindergarten through Third Grade Teachers by Year and Percent Present for One, Two, and Three Years for 18 Participating Schools.*

School	Year 1	Year 2	Year 3	Present one year	Present two years	Present three years
1	29	31	31	0.250	0.225	0.525
2	9	9	9	0.100	0.100	0.800
3	8	8	8	0.200	0.200	0.600
4	22	23	23	0.080	0.120	0.800
5	27	28	32	0.357	0.214	0.429
6	16	17	19	0.292	0.250	0.458
7	14	13	14	0.381	0.286	0.333
8	7	6	7	0.500	0.000	0.500
9	17	20	20	0.045	0.318	0.636
10	6	6	6	0.143	0.143	0.714
11	18	19	17	0.143	0.143	0.714
12	8	8	7	0.222	0.000	0.778
13	14	14	13	0.278	0.167	0.556
14	15	16	16	0.250	0.150	0.600
15	17	19	19	0.429	0.179	0.393
16	11	11	12	0.214	0.143	0.643
17	18	21	19	0.346	0.077	0.577
18	17	17	16	0.053	0.263	0.684
Total	273	286	288	0.251	0.183	0.566

Measures

We used a mix of reading assessments in order to assess broadly students' literacy learning over the primary grades in this study.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS). Participating students took a variety of subtests from DIBELS beginning in the fall of kindergarten through the fall of third grade. These subtests tap a range of early literacy skills, including letter recognition, phonological awareness, decoding, and oral reading fluency. The choice of sub-tests to administer at each grade level and occasion (fall and spring) was based primarily on publisher recommendations (Good & Kaminski, 2002). In some instances, however, we chose to include an additional, more difficult subtest, such as Oral Reading Fluency in the fall of first grade, in order to improve the capacity of our assessments to discriminate effectively among students with higher levels of literacy learning. Table 4 provides a schedule of the specific subtests administered at each grade-occasion. Reliability and validity of these subtests have been established and reported elsewhere (Good, Wallin, Simmons, Kame'enui, & Kaminski, 2002).

Terra Nova. Each spring, participating first through third grade students took the reading comprehension subtest from the Terra Nova Multiple Assessments of Reading, a group-administered, standardized, norm-referenced reading test. See McGraw-Hill (2001) for information on the reliability and validity of this test.

Table 4. *DIBELS and Terra Nova Testing Schedule by Grade and Time of Year.*

	Kindergarten		First Grade		Second grade		Third grade	
	Fall	Spring	Fall	Spring	Fall	Spring	Fall	Spring
DIBELS								
Initial sound fluency	X	X						
Letter name fluency	X	X	X					
Phonemic segmentation fluency	X	X	X	X				
Nonsense word fluency		X	X	X				
Oral reading fluency			X	X	X	X	X	
Terra Nova				X		X		X

Rasch scaling. The DIBELS and Terra Nova results were scaled together using Rasch modeling (Wright & Master, 1982). The resultant vertical scale allows us to exploit fully the longitudinal character of our student literacy learning data and solves some difficulties with the use of DIBELS assessments in program effects studies as described below.

First, DIBELS raw scores are often plagued with floor and ceiling effects since most of the subtests target constrained skills that are mastered in relatively short periods of time (Paris, 2005). The presence of such floor and ceiling effects can complicate analyses of program effects. By scaling various DIBELS sub-tests together in an integrated vertical metric, these problems are significantly reduced.

Second, the DIBELS and Terra Nova measure different aspects of reading development. The DIBELS subtests assess early literacy component skills relevant to the grade in which they are given. Since DIBELS was specifically designed to identify students at risk for subsequent reading failure, it discriminates best among students performing at lower literacy levels within each grade. In contrast, the Terra Nova does not focus on component skills but on comprehension, the ultimate goal of reading. As such, the Terra Nova discriminates more broadly and was seen as a good complement to the DIBELS subtests, which do not directly assess comprehension.

Thirdly and finally, the use of different tests at different times (see Table 4) made tracking literacy learning over time nearly impossible using only raw, or even publisher-scaled, scores. Moreover, raw test scores are problematic because the meaning of scoring one point more or less depends on the difficulty of the particular item sets composing the raw scores. In contrast, the Rasch scaling of test items calibrates the measure of children's reading abilities against the item difficulties associated with all of the tasks included in both the DIBELS and the Terra Nova assessments used in this study. The final vertical scale more closely approximates the principle of "a single ruler" or metric where a one-unit difference on the scale at any level of ability implies an equal difference on the trait measured (here, reading).

The full details of the Rasch rating scale analysis are reported elsewhere (Luppescu, Biancarosa, Kerbow, & Bryk, 2008). As is customary in Rasch scaling, the final measures are reported in a logit metric. Since logits are not intrinsically meaningful in terms of literacy skills, we illustrate here the differences in literacy status one would likely find among students scoring a 1, 2, 3, or 4 on our scale. A child scoring 1 logit (our average child in the fall of kindergarten), typically can name about 30 letters in a minute indicating good letter-name knowledge. That same child most likely discerns a few initial phonemes, but not many, and has very little chance of being able to segment words into phonemes. In contrast, a child scoring 2 logits (our average child in the spring of kindergarten and fall of first grade) is both accurate and fluent in letter name knowledge and has almost mastered initial sound identification, but is still largely unable to segment words phonemically. This same child can read a handful of words in a minute when given a passage of continuous text, but has little success at reading nonsense word (an indicator of "pure" decoding skill). A child scoring 3 logits (our average child in the spring of first grade and fall of second grade) has mastered letter names and initial sounds, can read 50-60 words per minute accurately, and may answer correctly about a third of the first grade Terra Nova comprehension questions. This child also does well on all but the hardest phonemic segmentation and nonsense word-reading tasks, but may not be very fast at these overall and is generally better at the former than the latter. Finally, a child scoring about 4 logits (our average child in the spring of second grade) has mastered component reading skills (e.g., letter name knowledge, phonemic segmentation, decoding), reads about 90 words correctly per minute, and does well on two-thirds of the first grade Terra Nova comprehension questions and on about a third of the second grade questions.

Table 5 reports the mean Rasch literacy development scores for K-2 students in the final analytic sample. Students gain about 1.2 logits during the academic year (from fall to spring) in kindergarten and first grade and about .90 logits during the second grade academic year. In contrast, the summer periods (i.e. between the spring and subsequent fall testing) are characterized by limited growth, or in some cases even a small loss, in literacy learning. The standard deviation in scores at all time periods is about 1 logit, except in the fall of second grade. Since only one DIBELS subtest (oral reading fluency) is used at this assessment point, scores appear more variable as a result of the larger standard errors of measurement resulting from this limited test battery.

Table 5. *Mean Scores (in Logits; n=22901) for Analytic Sample of Kindergarten (K) through Second Grade Students (n=7,961) by Grade and Time of Year.*

	K		First		Second	
	Fall	Spring	Fall	Spring	Fall	Spring
Mean	0.95	2.15	1.79	3.05	3.33	4.25
SD	1.39	1.37	1.23	0.94	1.99	1.19
N	3754	3826	3928	3940	3639	3814

Note: These results include data from all cohorts included in the study, both during the baseline period and first two years of LC program implementation.

Analyses and Results

We began our analyses by visually examining separately the observed mean outcomes for each cohort. Findings from this examination guided our approach to analyzing these data through hierarchical crossed-level value-added effects modeling.

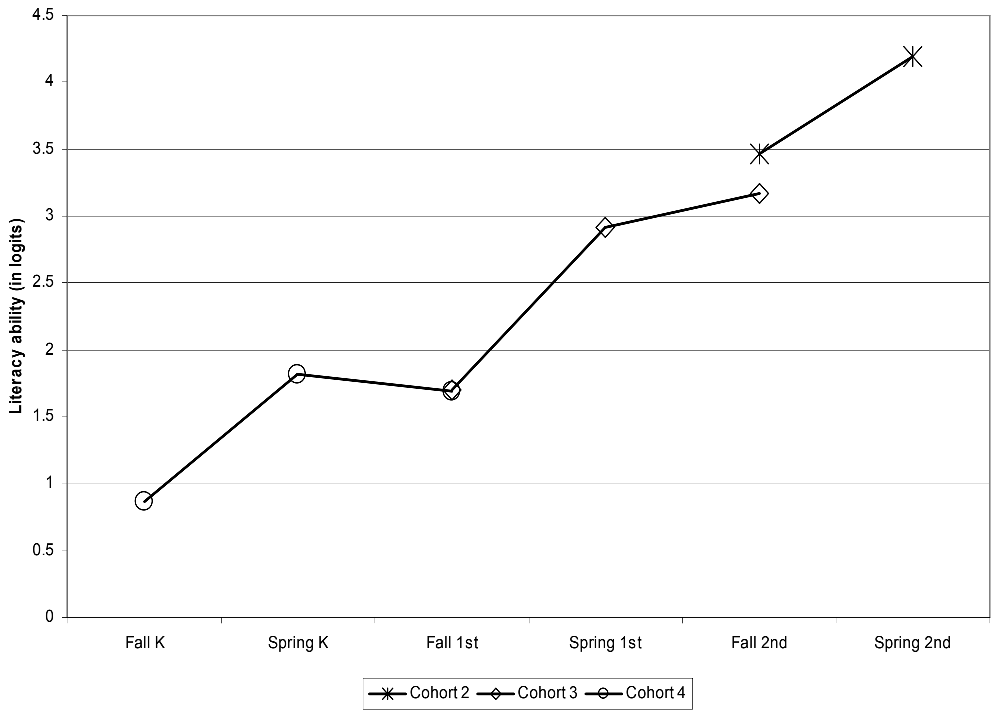
Empirical Student Literacy Learning Trajectories

We describe in this section the basic growth patterns found in the observed data. This examination revealed clear differences in learning rates by grade and during summer versus fall periods.

Baseline trend. Data collected in the fall and spring of the first year and fall of the second year (i.e. prior to initiation of LC school-based professional development) constitute the baseline trend for assessing subsequent program effects. Figure 3 depicts the observed means for Cohorts 2, 3, and 4 across these three testing occasions. Cohort 4 (represented by open circles in the figure) began the study in the fall of kindergarten, Cohort 3 (represented by open diamonds) in the fall of first grade, and Cohort 2 (represented by asterisks) in the fall of second grade. Given that the Literacy Collaborative model was not implemented in third grade and thus no value-added estimate can be assessed, the results from the third assessment occasion are not included for Cohort 2

Under an accelerated longitudinal cohort design, the results from the different baseline cohorts should interconnect smoothly together as one overall growth trajectory. The resultant longitudinal trajectory is the base against which subsequent program effects are evaluated. Note, as expected under an accelerated longitudinal cohort design, we found a near perfect overlap in mean achievement at the fall of first grade where Cohorts 3 and 4 join. A small gap of about 0.25 logits, however, was found where Cohorts 2 and 3 join in the fall of second grade. This indicates that prior to program implementation, these two cohorts differed somewhat in their average literacy ability, at least at this one time point. As a result, we have introduced a set of statistical adjustments for possible cohort differences in the hierarchical crossed-random effects models estimated below.

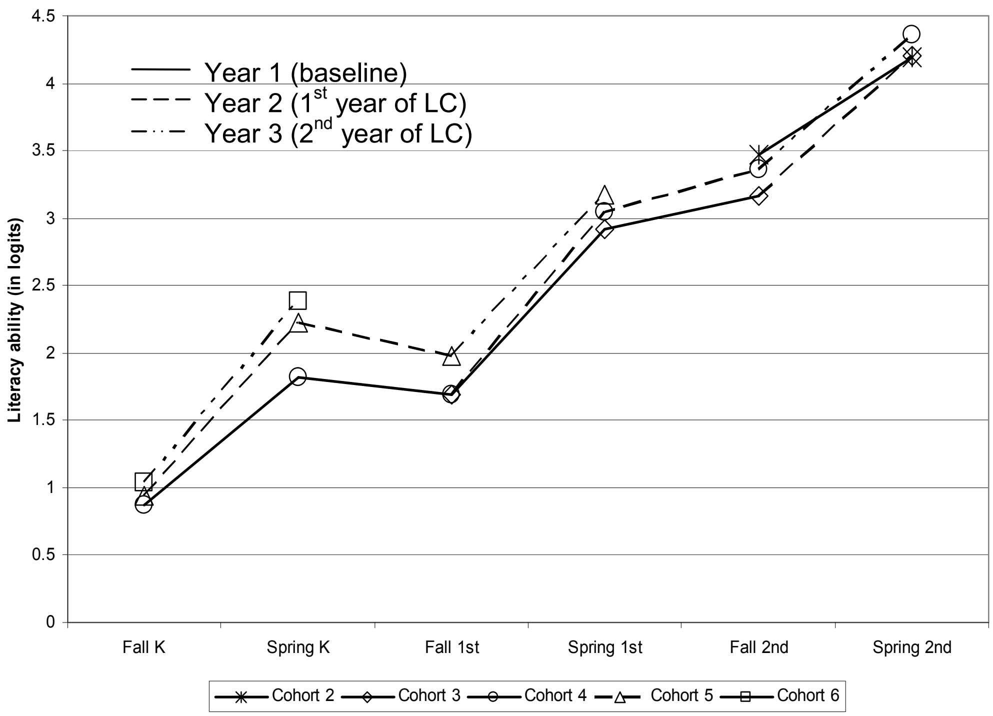
Figure 3. Means by cohort for the baseline period.



Implementation years. Next, we layered the subsequent LC implementation years' data on top of the baseline trend to provide a first look at possible program effects present in these data. We present in Figure 4 the mean student outcomes at each testing occasion, including during the first and second years of LC implementation. The longitudinal data for each separate cohort is identified by a distinct symbol. For example, the trajectory for Cohort 4 continues from the fall of first grade through the fall of second grade, representing first year LC implementation effects. Following the same cohort through to the spring of second grade incorporates second year LC implementation effects on this group.

Of primary interest in the figure is a comparison of the slopes representing student learning during the academic years and how these slopes change over the course of the study. Specifically, the increasing steepness of the slopes from fall to spring within each grade (from solid line, to dashed line, to dashed-and-dotted line) suggests possible overall value-added effects associated with the LC program. These value-added effects are most apparent in kindergarten where students' fall entry status is almost identical for Cohorts 4, 5 and 6, but there is increasing separation in achievement among these three groups by the following spring.

Figure 4. Means by cohort and year of Literacy Collaborative (LC) implementation.



Key observations for value-added modeling. In addition to the possible cohort effects in the baseline results noted earlier, several other distinct features in these longitudinal data have important implications for subsequent value-added modeling. First, growth during academic years (from fall to spring) is markedly steeper than growth during the summer periods (from spring to fall). This means that we must separately parameterize the rates of student learning in these two periods. Second, as noted earlier, the academic learning rates (slopes) appear to vary across grades levels with larger gains observed in kindergarten and first grade versus second grade.   Thus, we also need to introduce a set of fixed effects in the model to capture these departures from strict linearity.

Finally, there is some evidence in Figure 4 that program effects may vary by year of implementation and grade level. Thus, in the analyses below that assess teacher and school-level value-added effects to student literacy learning, we estimate separate effects for each year and grade.

Hierarchical Crossed-level Value-Added Modeling

The accelerated longitudinal cohort design used in the current study lends itself naturally to value-added modeling because our data consist of repeated measures on students who cross teachers within schools over time. The hierarchical crossed-level value-added effects model that we applied here can be conceptualized as the joining of two separate hierarchical models. One is a two-level model for individual growth in achievement over time, and the other is a two-level model of the value-added that each teacher and school contributes to student learning in each particular year. In essence, the core evidence for LC implementation effects consists of comparing learning gains in each teacher's classroom during program implementation to the gains in that same teacher's classroom during the base year. The observed gains in each classroom, however, are now adjusted for any differences over time in the latent growth trends for students being educated in each classroom. That is, unlike the simple descriptive statistics presented in Figures 3 and 4, a hierarchical crossed-level value-added effects model allows us to exploit fully the longitudinal character of the data on each student and adjusts for any outcome differences associated with individual latent growth trends in estimated teacher-classroom and school value-added effects.

Basic individual growth model. We began by specifying a level-1 model for the literacy score at time i for student j. Our individual growth model includes an intercept and two slope estimates, one for the academic year periods and one for summer periods, to account for observed differences in student learning between these time periods. In developing the time metrics at level 1, we sought to specify the intercept as the latent score for student i at entry into the data set, regardless of the specific time occasion when this may occur. We represent scores in subsequent time periods as an accumulation of subsequent academic year and summer learning rates for each child.

Specifically, we defined two coded variables to represent academic year and summer periods respectively. The academic time indicator incremented by one in the spring of each year while summer time incremented in the fall of each year (see Table 6). For example, for a Cohort 4 child who entered the data set in the fall of kindergarten, academic time would increment at the second, fourth, and sixth assessment waves, and summer time would increment at the third and fifth waves. Both indicators are set equal to zero when a student first enters the dataset regardless of the assessment wave. This preserves the meaning of the intercept as the initial reading status of the child upon entry into the study data collection.

The intercept and academic year learning rates were specified as randomly varying among individual students. These capture the variation between children in their initial literacy status and their growth in literacy during the academic year respectively. In preliminary analyses we considered modeling summer period effects as randomly varying as well. However, we were unable to reliably differentiate among children in this regard once random intercepts and random academic year learning effects were included in the model. Therefore, the summer period effect was treated as fixed at the individual child level.

Table 6. *Incrementing of Individual Growth Model Variables for Cohorts 2-6.*

		Kindergarten		First grade		Second grade	
Cohort	Variable	Fall	Spring	Fall	Spring	Fall	Spring
2	Intercept					1	1
	Academic_time					0	1
	Summer_time					0	0
3	Intercept			1	1	1	1
	Academic_time			0	1	1	2
	Summer_time			0	0	1	1
4	Intercept	1	1	1	1	1	1
	Academic_time	0	1	1	2	2	3
	Summer_time	0	0	1	1	2	2
5	Intercept	1	1	1	1		
	Academic_time	0	1	1	2		
	Summer_time	0	0	1	1		
6	Intercept	1	1				
	Academic_time	0	1				
	Summer_time	0	0				

i. Adjusting for possible time of entry effects. As noted earlier, a key feature of an accelerated multiple cohort design is that students enter the data set at different points in time by virtue of their cohort (i.e., Cohort 2 entered in second grade, Cohort 3 began in first grade, and Cohorts 4 through 6 entered in kindergarten). In addition, some students transfer into a school after the start of the study and are absorbed into their respective cohorts. To account for these facts, we entered another set of dichotomous indicator variables, termed "Location," as fixed effects in the child-level model. Each indicator variable corresponds to a specific grade and time of year and represents the first occasion at which a student appears in the dataset. For example, if a student joined the study in the fall of first grade year, that student's Location would be a three, and the Location_3 indicator variable scored as one and all other Location variables would be zero. The reference category for this set of indicators is fall of kindergarten, or Location_1. As a result, the intercept in the model represents the predicted initial literacy status in the fall of kindergarten and the Location variables for Locations two through six represent the mean differences in literacy status among students depending on when they first enter the dataset (see Appendix, Table A, for examples).

ii. Adjusting for cohort differences. Because the earlier descriptive analysis indicated small average differences in initial literacy status among some baseline cohorts, we also included a set of fixed effects for each cohort to absorb these and any other potential residual differences between cohorts (excluding Cohort 1, which entered the study in third grade, and Cohort 2, which served as a reference category). These dichotomous indicator variables were also included as fixed effects at the child level. We note that Cohort and Location effects are not redundant because students can enter a cohort at any time after data collection was initiated for their specific cohort. For instance, a Cohort 4 child who entered the study in the fall of first grade would have Location_3=1 and Cohort_4=1 (with other Location and Cohort indicators set to 0) whereas a child who entered at the onset of the study would have Location_3=0 and Cohort_4=1. Similarly, a Cohort 3 child would typically have Cohort_3=1 and Location_3=1 representing their entry in fall of first grade, but a transfer student who joined the study in the next year would have Cohort_3=1 and Location_5=1 (with other Location and Cohort indicators, including Location_3, set to 0). See Appendix, Table A, for further examples.

iii. Adjusting for possible measurement artifacts across grades. Finally, because the observed means did not follow a strict linear model across K-2, we added a final set of indicator variables, termed "Deviation," as fixed effects. These indicators were designated for assessments collected in the spring of first grade, the fall of second grade, and spring of second grade respectively. These capture the extent to which mean growth in the later grades is somewhat different than that observed in the earlier grade periods. See Table A in the Appendix for how the Deviation variables work in concert with the intercept and growth parameters, and Cohort and Location indicator variables.

The value-added model. The value-added model estimates both the average value-added by LC in each year of implementation and random value-added effects associated with each teacher and with each school in each year of implementation. We describe these below.

i. Average value-added effects. Separate fixed effects were estimated to assess the average value-added of LC during each year of program implementation. These estimates represent the increments above and beyond the underlying academic and summer learning trends in the baseline period. They assess the magnitude and significance of differences between the mean baseline trend and subsequent overall trends in achievement under LC implementation.

Specifically, we estimated separate average LC effects after the first year of implementation, the first summer after implementation, and after the second year of implementation. (These dichotomous indicator variables were termed LC_Year_1, LC_Summer, and LC_Year_2 respectively.) The LC_Year_1 indicator was set to one for observations collected in the spring of the second year of the study, LC_Summer was equal to one at the fall of the third year of the study, and LC_Year_2 was equal to one in the spring of the third year of the study, and all three were otherwise set to zero (see Appendix, Table B). These indicators, however, remain set at zero if this first occasion that the student appears in the data set.

Note the summer effect allows us to estimate the extent to which the value-added observed at the end of year 1 (spring testing) was maintained through the summer (i.e., the subsequent fall testing). If, in fact, the LC academic effect is sustained, then the estimate for the summer LC effect should be similar in magnitude to the first year LC effect.

ii. Random value-added effects for teachers and schools. The LC value-added effects in years 1 and 2 of implementation were allowed to vary randomly at the teacher and school levels. This allowed us to estimate the amount of variation among schools and among teachers within schools, and generate Empirical Bayes estimates of individual teacher and school effects.

In addition, the intercept was allowed to vary randomly among schools to represent potential selection effects associated with each school. In preliminary analyses, we also considered a random base effect for teachers within schools (i.e. mean differences among classrooms within schools in students' fall achievement during the base year) but found no reliable differentiation here. The academic slope was also allowed to vary randomly at the school level. This captures the variations among schools in academic year learning rates during the baseline period.

We also added a random effect, "Base_Tch_VA," to represent the variation among teachers within schools in their value added (or extra contribution) to student learning in the spring of the baseline year The dichotomous indicator variable, Base_Tch_VA, takes on a value of 1 only at this time point and only for those children who were also present in the data set in the fall of the baseline year. The fixed effect for Base_Tch_VA is set to zero so as not to be redundant with the academic year fixed effect.

Finally, the summer learning effects were allowed to vary randomly at the school level. Since there was no "assigned teacher" during the summer periods, a random teacher effect for the occasions was not deemed sensible.

Final model. Assembling all of these components together produced the following final model. The outcome, $Y_{ijk\ell}$, was defined as literacy status in logits in period *i* for student *j* in teacher *k*'s class in school ℓ. The first, or time-varying, level of the model was as follows:

$$Y_{ijk\ell} = \pi_{0jk\ell} + \pi_{1jk\ell}*(Academic\_time_{ijk\ell}) + \pi_{2jk\ell}*(Summer\_time_{ijk\ell}) + \pi_{3jk\ell}*(Deviation\_4_{ijk\ell}) + \pi_{4jk\ell}*(Deviation\_5_{ijk\ell}) + \pi_{5jk\ell}*(Deviation\_6_{ijk\ell}) + \pi_{6jk\ell}*(Base\_Tch\_VA_{ijk\ell}) + \pi_{7jk\ell}*(LC\_Year\_1_{ijk\ell}) + \pi_{8jk\ell}*(LC\_Summer_{ijk\ell}) + \pi_{9jk\ell}*(LC\_Year\_2_{ijk\ell}) + e_{ijk\ell}$$

The second or child-level model was defined as follows:

$$\pi_{0jk\ell} = \gamma_{00k\ell} + \gamma_{01k\ell}*(Location\_2) + \gamma_{02k\ell}*(Location\_3) + \gamma_{03k\ell}*(Location\_4) + \gamma_{04k\ell}*(Location\_5) + \gamma_{05k\ell}*(Location\_6) + \gamma_{06k\ell}*(Cohort\_3) + \gamma_{07k\ell}*(Cohort\_4) + \gamma_{08k\ell}*(Cohort\_5) + \gamma_{09k\ell}*(Cohort\_6) + r_{0jk\ell}$$

$$\pi_{1jk\ell} = \gamma_{10k\ell} + r_{1jk\ell}$$

$$\pi_{2jk\ell} = \gamma_{20k\ell}$$

$$\pi_{3jk\ell} = \gamma_{30k\ell}$$

$$\pi_{4jk\ell} = \gamma_{40k\ell}$$

$$\pi_{5jk\ell} = \gamma_{50k\ell}$$

$$\pi_{6jk\ell} = \gamma_{60k\ell}$$

$$\pi_{7jk\ell} = \gamma_{70k\ell}$$

$$\pi_{8jk\ell} = \gamma_{80k\ell}$$

$$\pi_{9jk\ell} = \gamma_{90k\ell}$$

The third level captured the effects of teachers on children as they cross classrooms within schools over time. It was expressed as:

$$
\begin{aligned}
\gamma_{00k\ell} &= \theta_{000\ell} \\
\gamma_{01k\ell} &= \theta_{010\ell} \\
\gamma_{02k\ell} &= \theta_{020\ell} \\
\gamma_{03k\ell} &= \theta_{030\ell} \\
\gamma_{04k\ell} &= \theta_{040\ell} \\
\gamma_{05k\ell} &= \theta_{050\ell} \\
\gamma_{06k\ell} &= \theta_{060\ell} \\
\gamma_{07k\ell} &= \theta_{070\ell} \\
\gamma_{08k\ell} &= \theta_{080\ell} \\
\gamma_{09k\ell} &= \theta_{090\ell} \\
\gamma_{10k\ell} &= \theta_{100\ell} \\
\gamma_{20k\ell} &= \theta_{200\ell} \\
\gamma_{30k\ell} &= \theta_{300\ell} \\
\gamma_{40k\ell} &= \theta_{400\ell} \\
\gamma_{50k\ell} &= \theta_{500\ell} \\
\gamma_{60k\ell} &= u_{60k\ell} \\
\gamma_{70k\ell} &= \theta_{700\ell} + u_{70k\ell} \\
\gamma_{80k\ell} &= \theta_{800\ell} \\
\gamma_{90k\ell} &= \theta_{900\ell} + u_{90k\ell}
\end{aligned}
$$

Note that this model includes three random teacher effects: a baseline year value-added, $u_{60k\ell}$, and first year and second year of implementation value-added effects ($u_{70k\ell}$ and $u_{90k\ell}$ respectively.)

Teachers were in turn nested in the fourth, or school-level which includes five random school effects— the differences among schools in initial status, $v_{000\ell}$, differences among schools in base year academic learning rates, $v_{100\ell}$, differences among schools in students' summer learning rates, $v_{200\ell}$. and their first and second year value-added effects, $v_{700\ell}$ and $v_{900\ell}$.

$$\theta_{000\ell} = \beta_{0000} + v_{000\ell}$$
$$\theta_{010\ell} = \beta_{0100}$$
$$\theta_{020\ell} = \beta_{0200}$$
$$\theta_{030\ell} = \beta_{0300}$$
$$\theta_{040\ell} = \beta_{0400}$$
$$\theta_{050\ell} = \beta_{0500}$$
$$\theta_{060\ell} = \beta_{0600}$$
$$\theta_{070\ell} = \beta_{0700}$$
$$\theta_{080\ell} = \beta_{0800}$$
$$\theta_{090\ell} = \beta_{0900}$$
$$\theta_{100\ell} = \beta_{1000} + v_{100\ell}$$
$$\theta_{200\ell} = \beta_{2000} + v_{200\ell}$$
$$\theta_{300\ell} = \beta_{3000}$$
$$\theta_{400\ell} = \beta_{4000}$$
$$\theta_{500\ell} = \beta_{5000}$$
$$\theta_{700\ell} = \beta_{7000} + v_{700\ell}$$
$$\theta_{800\ell} = \beta_{8000}$$
$$\theta_{900\ell} = \beta_{9000} + v_{900\ell}$$

The final mixed model was as follows:

$$\begin{aligned}
Y_{ijk\ell} = {} & \beta_{0000} + \beta_{01k\ell}*(Location_2) + \beta_{02k\ell}*(Location_3) + \beta_{03k\ell}*(Location_4) + \\
& \beta_{04k\ell}*(Location_5) + \beta_{05k\ell}*(Location_6) + \beta_{06k\ell}*(Cohort_3) + \beta_{07k\ell}*(Cohort_4) + \\
& \beta_{08k\ell}*(Cohort_5) + \beta_{09k\ell}*(Cohort_6) + r_{0jk\ell} + v_{000\ell} \\
& + \beta_{1000}*(Academic_time_{ijk\ell}) + r_{1jk\ell}*(Academic_time_{ijk\ell}) + v_{100\ell}*(Academic_time_{ijk\ell}) + \\
& u_{60k\ell}*(Base_Tch_VA_{ijk\ell}) \\
& + \beta_{2000}*(Summer_time_{ijk\ell}) + v_{200\ell}*(Summer_time_{ijk\ell}) \\
& + \beta_{3000}*(Deviation_4_{ijk\ell}) + \beta_{4000}*(Deviation_5_{ijk\ell}) + \beta_{5000}*(Deviation_6_{ijk\ell}) \\
& + \beta_{7000}*(LC_Year_1_{ijk\ell}) + u_{70k\ell}*(LC_Year_1_{ijk\ell}) + v_{700\ell}*(LC_Year_1_{ijk\ell}) \\
& + \beta_{8000}*(LC_Summer_{ijk\ell}) \\
& + \beta_{9000}*(LC_Year_2_{ijk\ell}) + u_{90k\ell}*(LC_Year_2_{ijk\ell}) + v_{900\ell}*(LC_Year_2_{ijk\ell}) \\
& + e_{ijk\ell}
\end{aligned}$$

Table 7 summarizes the interpretation for each of the parameters in the mixed model.

Table 7. *Interpretation of Coefficients in the Final Model.*

Variable	Name(s)	Coefficients	Variable Description
Intercept		β_{0000}	The base initial literacy status for children in the fall of kindergarten
		$r_{0jk\ell}$	The variance in initial literacy status between children nested within schools
		$v_{000\ell}$	The variance in initial literacy status between schools
Location, or point of entry	Location_2 .. Location_6	$\beta_{01k\ell}$ .. $\beta_{05k\ell}$	Each coefficient represents the mean difference in initial literacy status associated with entering the study other than in the fall of kindergarten (e.g., Location_2 indicates the difference in average initial literacy status for children who entered the study in spring of kindergarten rather than fall of kindergarten; Location_3 does so for children who entered the study in fall of first grade rather than fall of kindergarten and so on)
Cohort, or age group in accelerated cohort design	Cohort_3 .. Cohort_6	$\beta_{06k\ell}$ … $\beta_{09k\ell}$	Each coefficient represents the mean difference in initial literacy status for different cohorts of children in the accelerated cohort design compared to Cohort 2, which entered the study in the fall of second grade (Note: Location and Cohort effects should be evaluated jointly.)
Academic year growth parameter	Academic_time	β_{1000}	Baseline literacy learning rate during the academic year, or the average growth in literacy status from fall to spring, after adjusting for cohort, location, and value-added effects
		$r_{1jk\ell}$	Variance in latent academic year learning rates among students
		$v_{100\ell}$	Variance in school value-added effects on student learning during the baseline period
Random effect for teacher baseline year effects	Base_Tch_VA	$u_{60k\ell}$	Variance in teacher value-added effects on student learning (within schools) during the baseline period (i.e., from fall to spring in the first year of the study)

Summer months growth parameter	Summer_time	β_{2000}	Average literacy learning rate during the summer baseline period, adjusted for cohort and location.
		$v_{200\ell}$	Variance in summer literacy learning among schools
Adjustments for deviations from linear growth	Deviation_4 .. Deviation_6	β_{3000} .. β_{5000}	Adjustments to average growth trajectory to account for deviations from a constant growth rate across grades in academic and summer periods
First year of LC implementation	LC_Year_1	β_{7000}	Value-added effect for the first year of implementation averaged across schools and teachers
		$u_{70k\ell}$	Variance in the value-added effects for the first year of implementation among teachers within schools
		$v_{70k\ell}$	Variance in the value-added effects during the first year of implementation among schools
Summer between first and second years of implementation	LC_Summer	β_{8000}	Sustaining effect of the first year value-added through the summer, averaged across schools and teachers
Second year of LC implementation	LC_Year_2	β_{9000}	Value-added effect during the second year of implementation averaged across schools and teachers
		$u_{90k\ell}$	Variance in the value-added effects for the second year of implementation among teachers within schools
		$v_{90k\ell}$	Variance in the value-added effects for the second year of implementation among schools

Results

Estimates for the model above were derived using the HCM3 sub-program with HLM (beta version 7.0). HCM3 is a flexible four-level program that allows for estimating a variety of models. In our application, the model consisted of repeated measures (level 1) on students (level 2) crossing teachers (level 3) nested within schools (level 4). Formally we can represent this as (repeated measures: students) x (teachers: schools). All random effects at the teacher and school levels were treated as cumulative within HCM3. The full final fitted model is reported in Table C in the Appendix; only the most relevant results are discussed below.

Average value-added effects. As the results in Table 8 indicate, the latent growth rate for an average student in an average teacher's classroom in an average school was just over 1 logits in literacy learning during the baseline academic year. In terms of program effects, the average value-added during the first year of implementation was 0.17 logits. This represents a 16% increase in learning as compared to the average baseline growth rate. During the second year of implementation, the estimated value-added was 0.29, which represents a 27% increase in productivity over baseline growth. These value-added effects convert into standard effect sizes of 0.21 and 0.36 respectively based on the residual level-1 variance estimated under the HCM3 model. Interestingly, the baseline variation among students within classrooms in academic year growth, the baseline variation in value-added effects among teachers within schools, and among schools was about the same at 0.067, 0.069, and 0.060 respectively.

Table 8. *Average Value-added Effects and Variation in Effects among Students, Teachers, and Schools.*

	Fixed effects		Random effects		
	Coefficient (s.e.)	*t*	Variance among Students	Variance among Teachers within Schools	Variance among Schools
Baseline academic growth rate time	1.06*** (.07)	15.56	.067	.069	.060
Literacy Collaborative Value-added in Year 1	.17*** (.04)	4.45		.060	.012
Literacy Collaborative Value-added in Year 2	.29*** (.05)	5.93		.152	.012

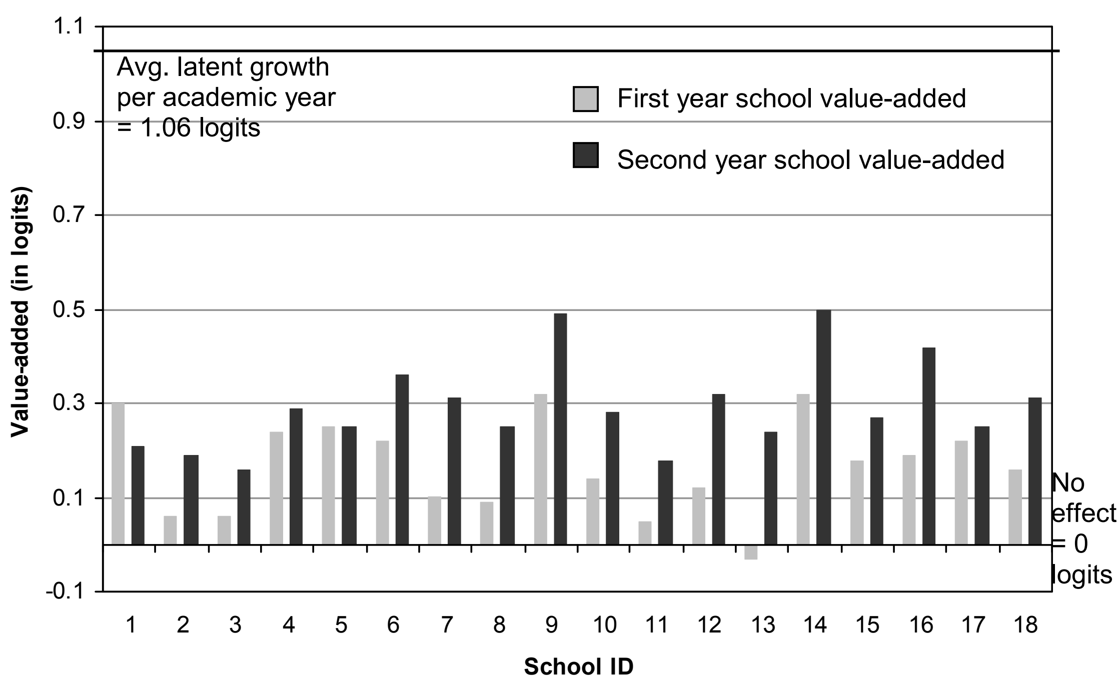
*** $p <= .001$

Variation in value-added effects. We also found significant variation in value-added effects among schools and among teachers within schools in both years 1 and 2. To better understand this inter- and intra-school variation, Figures 5 and 6 represent this variability based on the Empirical Bayes value-added estimates for each teacher and school in implementation years 1 and 2.

i. School level variance in value-added effects. As Figure 5 illustrates, most schools demonstrated a positive value-added in student literacy learning in both years of implementation. In the first year of implementation schools 1, 9, and 14 had the highest value-added estimates, about 0.3 logits, which represented a 28% increase in learning as compared to the baseline average learning rate of 1.06 logits per academic year. By the second year of implementation, two of these schools (9 and 14) demonstrated a value-added of about 0.5 logits, or a 47%

increase over the baseline rate. In contrast, several schools during the first year of implementation showed value-added effects of less than 0.1 logits, including schools 2, 3, 11, and 13. However, by the second year of implementation all of these schools showed value-added effects close to 0.2 logits or higher. In addition, it is notable that most schools demonstrated substantial increases in the value-added between the first and second years of implementation. The only exceptions are schools 1 and 5 but even here the estimate remained above 0.2 logits.[i] Finally, schools 7, 9, 12, 13, 14, and 16 demonstrated the largest observed increases in value-added (about 0.2 logits) from year one to year two of implementation.

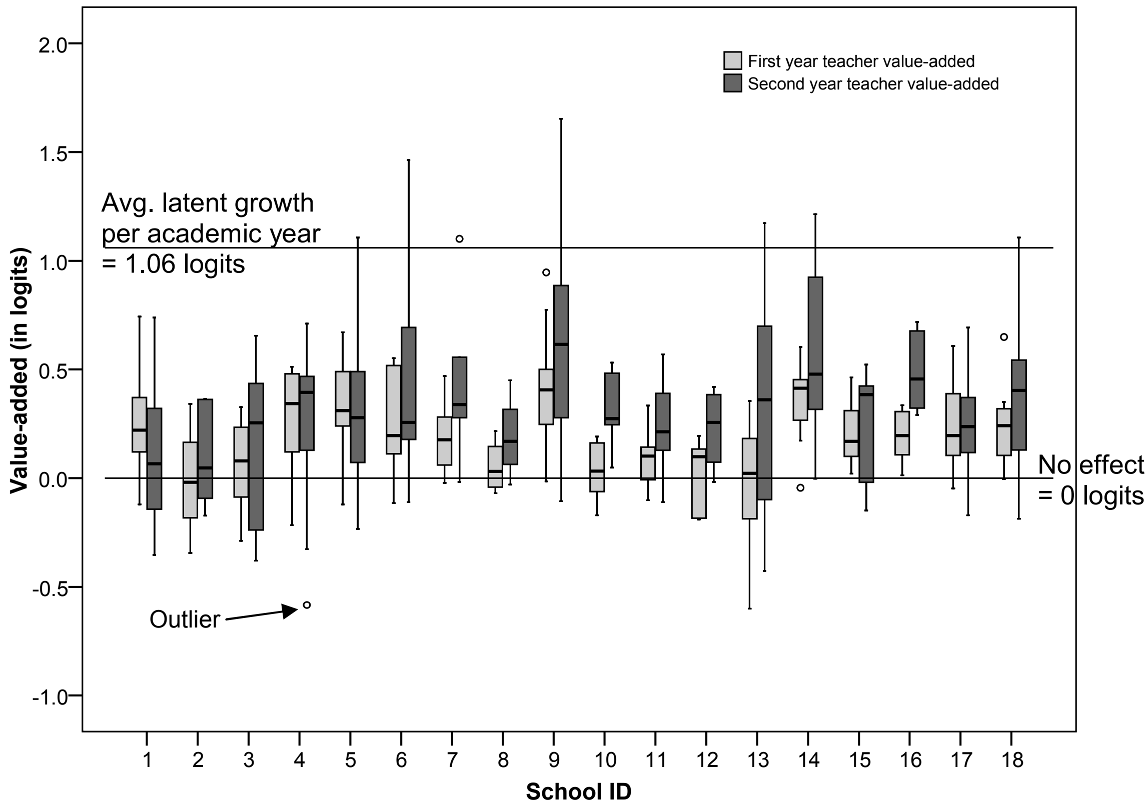
Figure 5. Variability in school value-added during two years of Literacy Collaborative implementation.



ii. Teacher-level variance in value-added effects. Figure 6 illustrates the variability in the Empirical Bayes estimates for the value-added effects associated with teachers within schools. The box-plots demonstrate considerable variability in effects among teachers within schools and that this variability increased in the second year of implementation relative to the first year. Of note, the two schools with the highest value-added estimates in both years of implementation (9 and 14) also showed a substantial increase in the variability among teacher effects, with at least some teachers in both schools demonstrating a value-added estimate of over 1.0 logits in the second year of implementation. However, even in these high value-added schools, some teachers still showed little to no program effect, as evidenced by the lower tails of the boxplots for these schools.

As might be expected, among the schools with the smallest value-added effects in the first year of implementation (2, 3, 11, and 13), some teachers showed positive effects, while others demonstrated negative value-added effects. However, by the second year of implementation, a majority of teachers even in these schools exhibited some positive value-added effects. Moreover, in school 13, at least one teacher showed a value-added of over 1.0 logits. Indeed, value-added estimates of over 1.0 logits were observed in several schools, including schools 5, 6, and 18. The largest overall shifts in the distribution of teacher value-added effects occurred in schools 7, 10, 11, and 16. In these schools, the 25th percentile for teacher value-added in the second year of implementation fell approximately at or above the 75th percentile for the first year of implementation.

Figure 6. Box-plots of variability in teacher value-added during the first and second year of Literacy Collaborative implementation.



Persistence of value-added effects over the summer. Although not a central question in the current study, the model also allowed us to examine whether the value-added effects estimated at the end of the first year of implementation were maintained over the subsequent summer. The results indicate that the first year effects did indeed persist through the subsequent fall testing. The average value-added effect after the first summer of LC implementation was 0.19 logits (t=5.953, p<.001), which is similar in magnitude to the first year implementation effect of 0.17 logits (t=4.446, p<.001).

Conclusion

Results from the current study demonstrate that Literacy Collaborative (LC) professional development is associated with a 16 to 29 percent improvement in student literacy learning rates in implementation years 1 and 2 respectively. Although these findings should be viewed as preliminary as we await analyses of data from the fourth and final year of the study, they do suggest substantial effects of LC professional development on student learning. Comparable results from the third year of implementation (i.e. the year 4 data collection) would further strengthen the empirical warrant for this conclusion.

Our results also indicate that the value-added by LC professional development varied considerably by school and by teacher within school. Of particular significance the variability among teacher effects appears to increase over time. Future analyses will investigate the extent to which school-level factors (including school size, amount of coaching, expertise of coaches, and school staff stability) and teacher-level factors (including experience teaching, amount of coaching, length of participation in LC professional development, and frequency and expertise of implementation of LC instructional practices) are predictive of these differences in the value-added to student learning.

Finally, results indicated persistence of first year effects over the following summer. Analyses based on the final year of data collection will allow us to examine whether estimated second year of implementation effects similarly persists through the following summer as well.

References

Allen, J. (2006). *Becoming a literacy leader: Supporting learning and change.* Portland, ME: Stenhouse Publishers.

Barton, R. & Lavrakas (2006). Finding gold at the end of the rainbow: Anchorage's investment in literacy coaching pays big dividends. *Northwest Regional Education Laboratory*, *12* (1), 6-11.

Bean, R.M. & Carroll, K.E. (2006). The literacy coach as a catalyst for change. In C. Cummins (Ed.), *Understanding and implementing Reading First initiatives: The changing role of administrators.* Newark, DE: International Reading Association.

Blachowicz, C. L. Z., Obrochta, C., & Fogelberg, E. (2005). Literacy coaching for change. *Educational Leadership*, *62* (6), 55-58.

Casey, K. (2006). *Literacy coaching: The Essentials.* Portsmouth, NH: Heinemann.

Clay, M. (1979). *Reading: The patterning of complex behavior* (2nd edition). Portsmouth, NH: Heinemann.

Clay, M. (2004). *Change over time in literacy learning*. Auckland, New Zealand: Heinemann Educational.

Fountas, I.C., & Pinnell, G.S. (1996). *Guided reading: Good first teaching for all children*. Portsmouth, NH: Heinemann.

Fountas, I.C. & Pinnell, G.S. (2006). *Teaching for comprehending and fluency: Thinking, talking, and writing about reading, K-8.* Portsmouth, NH: Heinemann.

Gibson, S. (2006). Lesson observation and feedback: The practice of an expert reading coach. *Reading Research and Instruction, 45,* 295-318.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Good, R. H., Wallin, J., Simmons, D. C., Kame'enui, E. J., & Kaminski, R. A. (2002). System-wide percentile ranks for DIBELS benchmark assessment (Technical report 9). Eugene, OR: University of Oregon.

International Reading Association (2004). *The role and qualifications of the reading coach in the United States*. Newark, DE: Author.

International Reading Association (2006). *Standards for Middle and High School Literacy Coaches*. Newark, DE: Author.

Luppescu, S., Biancarosa, G., Kerbow, D., & Bryk, A. S. (2008). "Testing the fluency-comprehension relationship through equating: A Rasch equating of DIBELS and Terra Nova in grades K-3." Society for Scientific Studies of Reading, Asheville, NC.

Marsh, J., Kerr, K., Ikemoto, G., Darilek, G., Suttorp, H., Zimmer, R., & Barney, H. (2005). *The role of districts in fostering instructional improvement: Lessons from three urban districts partnered with the Institute for Learning.* Santa Monica: Rand Corporation.

McGraw-Hill. (2001). *Terra Nova Technical Quality: Reliable, Useful Results Based on Psychometric Excellence*. Monterey, CA: CTB McGraw-Hill.

Neufeld, B. and Roper, D. (2003). *Coaching: A Strategy for Developing Instructional Capacity*. Cambridge: Education Matters, Inc.

Norton, J. (2007). Adding layers of support: Alabama's program helps site-based coaches succeed. *National Staff Development Council, 28,* 20-25.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40*, 184-202.

Poglinco, S. M., Bach, A. J., Hovde, K., Rosenblum, S., Saunders, M., & Supovitz, J. A. (2003). *The Heart of the Matter: The Coaching Model in America's Choice Schools*. Philadelphia: Consortium for Policy Research in Education.

Reed, B. & Rettig, R. (2006). Side by side: The Idaho Reading First program finds success by offering targeted professional development to its literacy coaches. *Northwest Regional Education Laboratory, 12* (1), 18-21.

Walpole, S. & McKenna, M.C. (2004). *The literacy coach's handbook: A guide to research-based practice.* New York: Guilford.

Wright, B. D., & Master, G. N. (1982). *Rating scale analysis*. Chicago: Mesa.

Appendix

This appendix includes two tables illustrating the flagging and incrementing of variables included in the models. An additional table reports the full final fitted hierarchical crossed-level value-added model.

Table A focuses on how the Cohort, Location, and Deviation variables work in concert with the intercept and growth rate parameters. Cohort and Location were dichotomous indicator variables included at the child level. Cohort captures the time occasion when a cohort first entered the study. Location captured the specific time period in which an individual child entered the study and was necessary because not every child in a cohort began the study with that cohort. For example, Cohort 4 children typically entered the study in the fall of their kindergarten year; however, some children were present for the spring testing but missed the fall testing, and others first joined the cohort in first or second grade. The Deviation indicator variables correct for deviations in student growth trajectories from a pure linear trend at specific assessment occasions (e.g. the fall of second grade).

Table B focuses on the value-added parameters for the first and second years LC implementation effects and on the year 1 summer effect. Table C reports coefficients, standard errors, *t* statistics, and *p*-values, as well as variances estimates and fit statistics for the final fitted model.

Table A. *Flagging and Incrementing of Individual Growth Model Variables and Adjustments Illustrated for Three Cohort 4 Children Entering the Study at Differing Times.*

		Kindergarten (K)		First grade		Second grade	
Time of entry	Variable	Fall	Spring	Fall	Spring	Fall	Spring
K fall	Intercept	1	1	1	1	1	1
	Academic_time	0	1	1	2	2	3
	Summer_time	0	0	1	1	2	2
	Location_2	0	0	0	0	0	0
	Location_3	0	0	0	0	0	0
	Location_4	0	0	0	0	0	0
	Location_5	0	0	0	0	0	0
	Location_6	0	0	0	0	0	0
	Cohort_3	0	0	0	0	0	0
	Cohort_4	1	1	1	1	1	1
	Cohort_5	0	0	0	0	0	0
	Cohort_6	0	0	0	0	0	0
	Deviation_4	0	0	0	1	0	0
	Deviation_5	0	0	0	0	1	0
	Deviation_6	0	0	0	0	0	1
K spring	Intercept		1	1	1	1	1
	Academic_time		0	0	1	1	2
	Summer_time		0	1	1	2	2
	Location_2		1	1	1	1	1
	Location_3		0	0	0	0	0
	Location_4		0	0	0	0	0
	Location_5		0	0	0	0	0
	Location_6		0	0	0	0	0

Time of entry	Variable	Kindergarten (K)		First grade		Second grade	
		Fall	Spring	Fall	Spring	Fall	Spring
	Cohort_3		0	0	0	0	0
	Cohort_4		1	1	1	1	1
	Cohort_5		0	0	0	0	0
	Cohort_6		0	0	0	0	0
	Deviation_4		0	0	1	0	0
	Deviation_5		0	0	0	1	0
	Deviation_6		0	0	0	0	1
1 fall	Intercept			1	1	1	1
	Academic_time			0	1	1	2
	Summer_time			0	0	1	1
	Location_2			0	0	0	0
	Location_3			1	1	1	1
	Location_4			0	0	0	0
	Location_5			0	0	0	0
	Location_6			0	0	0	0
	Cohort_3			0	0	0	0
	Cohort_4			1	1	1	1
	Cohort_5			0	0	0	0
	Cohort_6			0	0	0	0
	Deviation_4			0	1	0	0
	Deviation_5			0	0	1	0
	Deviation_6			0	0	0	1

Table B. *Flagging of Average Value-added Effects Relative to Individual Growth Model Variables for Three Cohorts.*

		Kindergarten		First grade		Second grade	
Cohort	Variable	Fall	Spring	Fall	Spring	Fall	Spring
3	Intercept			1	1	1	1
	Academic_time			0	1	1	2
	Summer_time			0	0	1	1
	LC_Year_1			0	0	0	1
	LC_Summer			0	0	0	0
	LC_Year_2			0	0	0	0
4	Intercept	1	1	1	1	1	1
	Academic_time	0	1	1	2	2	3
	Summer_time	0	0	1	1	2	2
	LC_Year_1	0	0	0	1	0	0
	LC_Summer	0	0	0	0	1	0
	LC_Year_2	0	0	0	0	0	1
5	Intercept	1	1	1	1		
	Academic_time	0	1	1	2		
	Summer_time	0	0	1	1		
	LC_Year_1	0	1	0	0		
	LC_Summer	0	0	1	0		
	LC_Year_2	0	0	0	1		

Table C. *Final Hierarchical Crossed-level Value-added Model, including Fixed and Random Effects.*

Fixed Effects	Coefficient	S.E.	*t*	p-value
Intercept	1.20	0.14	8.59	0.000
Location_2	0.93	0.08	12.05	0.000
Location_3	0.82	0.05	14.37	0.000
Location_4	1.83	0.10	18.26	0.000
Location_5	1.61	0.09	17.17	0.000
Location_6	2.54	0.13	19.26	0.000
Cohort_3	-0.33	0.08	-4.31	0.001
Cohort_4	-0.36	0.08	-4.50	0.001
Cohort_5	-0.26	0.09	-2.98	0.011
Cohort_6	-0.18	0.09	-1.92	0.077
Academic_time	1.06	0.07	15.56	0.000
Summer_time	-0.19	0.07	-2.79	0.016
Deviation_4	0.06	0.04	1.47	0.165
Deviation_5	0.64	0.05	13.54	0.000
Deviation_6	0.35	0.08	4.65	0.000
LC_Year_1	0.17	0.04	4.45	0.001
LC_Year_2	0.29	0.05	5.93	0.000
LC_Summer	0.19	0.03	5.95	0.000

Random Effects	Variance
Residual	0.55564
Student level	

Intercept	1.41340
Academic_time	0.06706
Teacher level	
Base_AY	0.06943
LC_Year_1	0.06032
LC_Year_2	0.15235
School level	
Intercept	0.21866
Academic_time	0.06015
Summer_time	0.06837
LC_Year_1	0.01201
LC_Year_2	0.01217

Fit	
Deviance	68122.264
Parameters	43

[i] This school had substantial teacher turnover during the course of the study. Future analyses will investigate the extent to which teacher turnover and stability may contribute to the value-added during implementation.