INVESTIGATING THE EFFECTIVENESS OF A COMPREHENSIVE LITERACY COACHING PROGRAM IN SCHOOLS WITH HIGH TEACHER MOBILITY

ABSTRACT

Teacher mobility is a factor that impacts schoolwide implementation of professional development programs. In this article, we present interim results of a longitudinal randomized field trial of a comprehensive literacy coaching program (Content-Focused Coaching, CFC) for improving instruction and learning in schools with high teacher mobility. We investigate program effects on 73 new treatment and comparison teachers recruited to replace the large proportion of teachers who left their schools during the first year of the program. HLM analyses indicated that the CFC program predicted significantly higher school-level gains on the state standardized test for English language learners ($N = 496$, $ES = .51$). By spring, the quality of teachers' self-reported and observed instruction in the CFC schools exceeded that of comparison teachers. Implications for accommodating new teachers into an ongoing and established coaching program to improve instruction and student learning, and conducting randomized trials in schools with high teacher turnover, are discussed.

Lindsay Clare Matsumura
UNIVERSITY OF PITTSBURGH

Helen E. Garnier
LRDC/UNIVERSITY OF PITTSBURGH
BIOLOGICAL SCIENCES CURRICULUM STUDIES, COLORADO SPRINGS

Richard Correnti
LRDC/UNIVERSITY OF PITTSBURGH

Brian Junker
CARNEGIE MELLON UNIVERSITY

Donna DiPrima Bickel
LRDC/UNIVERSITY OF PITTSBURGH

THE ELEMENTARY SCHOOL JOURNAL VOLUME 111, NUMBER 1

LITERACY coaching has been adopted by nearly every urban school district in the country as a strategy for increasing the reading achievement of poor, minority, and English language learning (ELL) students. Moreover, coaching is a key component included in state and federal literacy reform initiatives. Despite the widespread endorsement of literacy coaching, however, the evidence that coaching is an effective strategy for improving instruction and learning remains relatively weak. Most of the past research limited investigating the effectiveness of coaching to qualitative and nonexperimental designs (e.g., Joyce & Showers, 1996) or framed the impact of coaching as part of a larger set of reform activities (e.g., Camburn, Kimball, & Lowenhaupt, 2008). Few studies have applied controlled designs to investigate coaching's specific effect on teaching and learning.

More recent studies funded by the Institute of Education Sciences (IES) and other organizations aim to address this gap in coaching research. In the current article, we report findings from one such study—a longitudinal randomized field trial investigating the effect of a professional development (PD) program for coaches (Content-Focused Coaching, CFC)—on reading comprehension instruction and students' reading achievement in an urban district. Specifically, we investigated the effect of the CFC program on teaching and learning for teachers who were recruited midway through the program's implementation. These entering teachers replaced the large number of established teachers who left their school or grade between the first and second year of the trial (nearly half the sample). Our goal was to understand the effect of CFC on teachers who participated in the program for variable lengths of time and generate information that could be used to support the design and optimal implementation of comprehensive literacy coaching programs in schools that experience high rates of teacher mobility.

Teacher Mobility

High teacher turnover is common in urban districts. According to results from the Teacher Follow-Up Survey (Marvel, Lyter, Peltola, Strizek, & Morton, 2006), about 16% of public school teachers leave their school during any given year. In schools serving high numbers of low-income and minority students, teacher mobility rates are much higher (Henusheck, Kain, & Rivkin, 2004; Scafidi, Sjoquist, & Stinebrickner, 2005). Research conducted in the Chicago public schools, for example, shows that schools that serve low-income and primarily African American and Latino students lose a quarter or more of their teachers each year. The typical elementary school in the district loses half its teaching staff within 5 years, and many schools lose half their staff within 3 years (Allensworth, Ponisciak, & Mazzeo, 2009). The primary reasons given by teachers for leaving hard-to-staff schools are poor relations with the parents and student disciplinary problems (Allensworth et al., 2009).

High teacher mobility creates numerous problems for schools. First, the new replacement teachers tend to be among the least experienced and least qualified

teachers in the school (Reichardt, 2008). School leaders (principals and teachers) must devote a great deal of time to mentoring new teachers in order to ensure that they attain at least a minimum level of competency. This pattern is troubling since the large majority of schools with high teacher mobility tend to serve low-income students with the greatest learning needs (Reichardt, 2008). More worrisome are research findings that more effective teachers tend to move to schools with fewer low-income, minority, and lower-achieving students, which pose less teaching challenges (Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2008; Reichardt, 2008). The less effective teachers who leave their school, in contrast, tend to move to similarly low-performing schools (Boyd et al., 2008).

Second, teacher mobility presents a substantive challenge to sustaining and deepening instructional reform programs within a school. In addition to diverting school leaders' attention away from reform efforts, an influx of new teachers each year can disrupt existing collaborative relationships between teachers and undermine efforts to create opportunities in schools for teachers to improve their practice across academic years. Faculty discontinuity is of particular concern given research linking sustained, intensive, and collaborative professional learning opportunities for teachers to improved instruction and learning (Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009).

Research on the Effectiveness of Coaching

The limited research that exists on the effectiveness of coaching generally indicates positive effects on teachers' practice. A series of studies conducted by Joyce and Showers in the 1980s found that teachers who participated in peer coaching were more likely to apply new strategies in their teaching (Joyce & Showers, 1996). Neufeld and Roper (2003) and Knight (2004) similarly found that teachers actively involved in coaching tried new instructional practices learned in traditional workshops more often than teachers who did not participate in coaching.

Results on the effectiveness of coaching for improving student achievement are mixed. On the positive side, Marsh et al. (2008) found that coaching exerted a small but positive effect on reading achievement in two of the four cohorts of students included in their study of a statewide middle school reading coach program. Marsh et al. (2008) also found a small, significant relationship between how often coaches reviewed assessment data with teachers and student achievement. Sailors and Price (2010) found that elementary school teachers who received coaching in addition to participating in a 2-day workshop scored higher on all measures of instruction and student learning than teachers who only participated in the 2-day workshop. Interim results from the study of the Literacy Collaborative (LC)—a PD program for coaches—similarly found that teachers' participation in coaching had a positive effect on student achievement (Biancarosa, Bryk, & Dexter, 2008).

On the other hand, no effects for student achievement appeared in a national evaluation of the Reading First initiative that includes a substantial focus on coaching (Gamse, Jacob, Horst, Boulay, & Unlu, 2008) or a recent study con-

ducted by Garet et al. (2008). In the Garet et al. (2008) study, elementary schools were randomly assigned to one of three conditions: to participate in PD institutes aimed at increasing teachers' knowledge of scientifically based reading instruction, to participate in those same institutes while also receiving coaching, or to serve as a control group. A positive effect appeared for both of the PD interventions on teachers' knowledge of scientifically based reading instruction and on teachers' observed instruction, but not for student achievement (Garet et al., 2008). No added benefit of coaching was detected on the teachers' practice relative to teachers having only participated in the PD institutes.

Overview of the Study and Research Questions

To study the CFC program's effectiveness, elementary schools that served high numbers of low-income, minority, and ELL students were randomly assigned to participate in the CFC program or to serve as a comparison sample (i.e., to continue with the PD resources that were standard in the district, which included the use of literacy coaches). Data on fourth- and fifth-grade teachers' participation in literacy coaching, self-reported and observed instruction, and students' reading achievement were collected over a 3-year period.

Analyses we conducted from a previous study focused on the teachers ($N = 98$) who remained in their schools for the first 2 years of the study (referred to as Cohort 1 teachers) and their students. We found that after 2 years, Cohort 1 teachers in the CFC schools significantly improved the quality of their observed classroom text discussions relative to the teachers in the comparison schools, and the school-level achievement of their ELL students ($N = 741$) was significantly higher than that of their comparison peers on the state standardized achievement test, but not on the Degrees of Reading Power assessment (Matsumura, Garnier et al., 2009).

Our plan for studying the effectiveness of the CFC program was to follow coaches, principals, and teachers over a 3-year period. We anticipated teacher turnover during the study, however, nearly half the teacher sample, a much larger proportion than expected, left their school or grade between the first and second year of the trial. Since the CFC coaches were intended to work with all eligible teachers in a school and engage these teachers in professional learning communities, we decided to recruit the teachers who had been hired to replace those teachers who had left their school or targeted grade. In the current article, we focus on the teachers who were hired (referred to as Cohort 2 teachers) to replace those who had left the school. As described earlier, our goal was to extend our previous work by investigating the CFC program's effect on teachers who entered the study at different stages of the program and participated for varying lengths of time in the hopes of generating information that could be used to support the design, and maximize effective implementation, or comprehensive literacy coaching programs in schools with high teacher mobility.

The specific research questions the current article addresses are as follows: (1) What is the effect of a school's participation in the CFC program on Cohort 2

teachers' experience of literacy coaching: participation in coaching, perception of the usefulness of coaching, and the content emphasized in the coaching activities in which teachers participated? (2) Do Cohort 2 teachers in schools participating in the CFC program improve the quality of their reading comprehension instruction? and (3) Do students, particularly ELL students of Cohort 2 teachers in schools participating in the CFC program, improve in their reading comprehension skills?

Description of the CFC Program

The CFC program was originally developed for use in mathematics (West & Staub, 2003). The program was adapted to support literacy instruction and learning at the University of Pittsburgh's Institute for Learning (IFL; Staub & Bickel, 2003). Literacy coaches engaged in 3 days of PD a month led by fellows from the IFL (Donna DiPrima Bickel and Kathleen McCarthy). The goal of the PD program was to develop coaches' knowledge of effective reading comprehension instruction and pedagogical expertise and skill at working effectively with teachers. Principals and district leaders also participated in the sessions to help create the conditions in schools that would support effective coaching. The pedagogical focus of the CFC program and the strategies used by the IFL fellows to build the participants' knowledge and skills are described in the following sections.

Pedagogical Focus

The CFC training focuses on developing coaches' knowledge of the theory underlying reading comprehension instruction, with a special emphasis on the role of classroom talk in supporting students' understanding of texts (August & Shanahan, 2006; Beck & McKeown, 2006; Goldenberg, 2008; Snow, 2002). Research shows that the reading skills of low-income children decline steeply in the upper elementary grades when the focus of reading shifts from decoding to higher-level comprehension (Chall & Jacobs, 2003). A key strategy for promoting students' higher-level comprehension skills is to engage students in authentic classroom discussions that provide opportunities for both conceptual and linguistic development (Tharp & Gallimore, 1988). Specifically, class discussions that build upon students' own understanding and experiences while encouraging students to express their ideas have been shown to promote deeper comprehension of a text's meaning and develop students' higher-level cognitive skills. This is true as well for ELL students, who comprised a significant percentage of the students in the schools in which the intervention was conducted (Goldenberg, 2008). For example, a recent report by the National Literacy Panel on language-minority children and youth found that direct instruction on reading fundamentals (e.g., letter-sound relationships) in tandem with the development of "more thorough discussion routines around literature" that provide students with multiple opportunities for language use is the most effective path for improving ELL students' reading comprehension skills (Shanahan & Beck, 2006, p. 448).

To improve the quality of classroom talk in schools, CFC coaches studied and applied techniques from *Questioning the Author* (Beck & McKeown, 2006; Beck, McKeown, Hamilton, & Kucan, 1997). Questioning the Author is designed to help teachers choose texts that support meaningful discussions, identify the major ideas students should construct from a text, anticipate potential problems students might have comprehending a text, segment a text to focus on information students might need to build comprehension, and develop interpretive questions that promote and deepen comprehension. Coaches studied Questioning the Author with other coaches and the IFL fellows in preparation for studying the text with teachers at their school.

Coaches then worked intensively with the IFL fellows to learn how to plan and model Questioning the Author lessons in teachers' classrooms. Drawing on principles of cognitive apprenticeship and practice-based PD (Ball & Cohen, 1999; Collins, Brown, & Holum, 1991), the coaches studied models of instruction and enacted these instructional practices with assistance before moving on to independent enactment in teachers' classrooms. For example, the IFL fellows modeled Questioning the Author lessons for the coaches in teachers' classrooms before the coaches then practiced teaching Questioning the Author lessons in teachers' classrooms while being observed by the IFL fellows and their fellow coach trainees. Coaches were provided with multiple opportunities to practice teaching reading comprehension lessons before moving on to independently model lessons in teachers' classrooms.

Coaching Expertise

In addition to building coaches' pedagogical knowledge and skills, the CFC training focuses on building coaching skills, that is, coaches' ability to work with teachers productively in their classrooms and in school-based professional learning communities (McLaughlin & Talbert, 2006). To accomplish these goals, coaches engaged in labs in which they first observed the IFL fellows coaching teachers, and later were observed working with teachers at their school individually and in grade-level teams. Over the course of the year, all CFC coaches were provided with the opportunity to observe other coaches as well as be observed and receive feedback on their coaching practices.

Engaging School and District Leaders

Principals and central-office leaders also participated in the PD sessions for coaches one day a month. The purpose of engaging school leaders in the coach training is to educate them about concrete ways in which they can support coaches' work with teachers and help establish a common understanding of the responsibilities of a literacy coach that are shared across the levels of the school system. This latter goal is especially important given that principals are coaches' immediate supervisors, including the CFC coaches in our study. Principals' understanding and endorsement of the program is essential for ensuring that the CFC coaches are permitted to focus their time on working directly with teachers

to improve their practice instead of on other tasks often assigned to literacy coaches (e.g., tutoring students, coordinating assessments, etc.).

The CFC program developers expected coaches to meet with teachers in weekly grade-level team meetings to study the theories underlying effective reading comprehension instruction and to plan Questioning the Author lessons. Coaches also were expected to meet individually with teachers on a monthly basis to engage in a cycle of planning, enacting Questioning the Author lessons (i.e., modeling lessons or observing teachers enact these lessons), and reflecting on instruction. The purpose of these activities was to model new ways of thinking about instruction and to provide direct assistance to teachers for improving the quality of their reading comprehension instruction directed at the quality of classroom text discussions. Higher-quality classroom text discussions, in turn, were expected to improve students' reading comprehension skills.

Method

Setting and Participants

The randomized trial was located in a medium-sized urban district in Texas. Thirty-two elementary schools serving the lowest-achieving students in the district were randomly assigned treatment and comparison conditions. The assignment of schools to condition took place in a public process presided over by the research staff and district leadership.[1] Prior to the beginning of the CFC intervention, one comparison school left the trial, resulting in a final sample of 29 schools (15 treatment schools and 14 comparison schools).

Coaches. For the purpose of the randomized trial, school district leaders and fellows from the University of Pittsburgh's Institute for Learning collaborated to hire the CFC coaches. The hiring procedure required coach candidates to submit samples of their lesson plans and comment on a videotaped reading comprehension lesson. The coach candidates also shared discussion questions they might use in leading a study of a professional text with teachers. Eleven CFC coaches were hired in the first year of the trial (some coaches were shared across two smaller schools) and an additional four coaches were hired in the trial's second year ($N = 15$). Ten of the 14 comparison schools also had literacy coaches.

CFC coaches had an average of 12 years of elementary teaching experience (ranging from 0 to 40 years) and 2 years of prior coaching experience (ranging from 0 to 22 years). The one coach who did not have prior classroom teaching experience held a doctorate in psychology and had worked for 8 years at a large education research center investigating the connections between schoolwide reform efforts, literacy instruction, and students' reading outcomes, and conducting PD for literacy coaches. Four other CFC coaches held a master's degree.

Literacy coaches in the comparison schools who were working with the upper elementary grade teachers, but were not Reading First coaches, were also recruited to participate in the trial. During the second year of the trial, 10 of the 14 comparison schools had literacy coaches. The coaches had an average of 15 years of prior teaching experience (ranging from 4 to 40 years) and 3 years of prior coaching

experience (ranging from 1 to 10 years). More than half of the coaches (60%) held graduate degrees. CFC and comparison coaches did not differ significantly on measures of education and experience.

Teachers. Initially, 193 fourth- and fifth-grade teachers in the study schools agreed to participate in the trial and 177 continued to participate through year 1. The primary reason teachers left the trial in year 1 was to take maternity leave or personal leave. Compared to teachers who remained in the trial, those who left differed significantly on only one of the above measures; they reported less experience teaching reading (4 years compared to 8 years, on average). Within the group of teachers remaining in the study, comparison and intervention teachers did not differ on years experience teaching reading. Of these teachers (Cohort 1 teachers), 98 continued to participate through year 2. Comparisons between teachers who stayed and teachers who were no longer in the trial at the beginning of year 2 detected no differences in education, teaching experience, or certification backgrounds, with one exception: 73% of teachers who left spoke only English in their classroom, compared to 53% of the teachers who remained in the trial ($p < .05$). Of the teachers who left the trial, both the CFC and comparison groups reported similar background characteristics. The rate of teacher attrition did not differ between the treatment and comparison groups.

At the beginning of year 2, we recruited 73 fourth- and fifth-grade teachers (Cohort 2) to replace those who left the trial. Fifty-eight percent of the Cohort 2 teachers taught in the treatment (CFC) schools and 42% taught in the comparison schools. No background differences were detected between the Cohort 2 teachers in the treatment and comparison conditions. Upon entry, all of the Cohort 2 teachers taught full-time: 73% taught fourth grade and 27% taught fifth grade. Over a third (37%) entered with master's degrees and 100% were certified to teach. Sixty percent taught a designated ELL class, 6% taught an honors or gifted and talented class, and 4% taught a designated special education class. As a group, these teachers had a wide range of experience (0 to 35 years) and averaged about 6 years of teaching reading. English was the language of instruction for the majority of teachers (56%), while approximately 38% taught in both English and Spanish, and 4% taught in Spanish only.

Both teacher cohorts were similar, with two exceptions. First, more Cohort 2 teachers taught fourth grade (73%) than Cohort 1 teachers (58%). Second, Cohort 2 teachers had fewer years of experience teaching at their current grade level (6 years) than Cohort 1 teachers (9 years).

Students. In the spring of 2008, the school district provided complete student demographic information and reading achievement scores for 1,269 students (70% fourth grade, 30% fifth grade) of the Cohort 2 teachers. The large majority of these students qualified for free or reduced-price lunch (91%). Students were mostly Hispanic (81%) or African American (13%). The remaining students were white or Asian (2% and 4%, respectively). Forty percent of the students were designated as ELL students by the school district. CFC schools included more fourth-grade students (74%), ELL students (43%), and students of Asian ethnicity (3%), while comparison schools included more African American students (17%). These differences were included as covariates in the analyses.

Measures and Procedures

Teacher surveys. In each trial year of the study, the same measures were collected and the same procedures were used to collect them. Teachers were administered surveys at the beginning (August) and end (April) of each academic year. Teachers described their prior education and teaching experience, their frequency of participating in different literacy coaching activities, the content emphasized in the coaching activities, the usefulness of coaching for improving their practice, the quality of classroom text discussions, and the quality of their school's professional community.

Teacher participation in coaching. The frequency of teachers' participation in literacy coaching activities emphasized in the CFC program was assessed on five items (coded as 1 = never, 2 = one to three times a year, 3 = four to six times a year, 4 = monthly, 5 = weekly): (*a*) coach met with me and other teachers in grade-level meetings, (*b*) coach met with me individually, (*c*) coach observed me teaching for at least 30 minutes, (*d*) coach taught a model lesson for me, and (*e*) coach taught a lesson with me in my classroom. These items were summed to create one measure of the teacher's overall level of participation in the five activities ($\alpha = .85$).

Teachers also described the usefulness of the participation in each of these activities for improving their practice (coded as 1 = not useful to 4 = very useful). The five items were averaged to create one measure of perceived usefulness of coaching activities ($\alpha = .90$).

Content of coaching activities. Teachers' perceptions of the content emphasized in the coaching activities were assessed with 16 items (coded as 1 = no emphasis to 4 = major emphasis) and averaged into four constructs ($\alpha = .81$–$.89$): (*a*) building knowledge of the theory underlying effective reading comprehension instruction (e.g., coaches study education research with teachers), (*b*) planning and reflecting on instruction (e.g., coaches help teachers plan rigorous instruction, reflect on their instruction after a lesson, and work with teachers to better anticipate student difficulties or misconceptions), (*c*) providing help during lesson enactment (e.g., coaches help teachers facilitate better classroom discussions and teach new instructional strategies), and (*d*) differentiating instruction (e.g., coaches help teachers plan lessons tailored to specific student learning needs).

Reading comprehension instruction. Teachers described the typical occurrence in their instruction of 11 different activities related to student discussions (coded 1 = never to 4 = almost always). Averages were created to measure two constructs ($\alpha = .70$ and $.78$): (*a*) students communicate with each other in class discussions (e.g., students build on each other's ideas in class discussions, students provide constructive feedback to their peers), and (*b*) students make connections during discussions (e.g., students draw on relevant academic knowledge to support their assertions, make connections across texts, and connect ideas and events within a text).

Classroom observations. Raters used the Instructional Quality Assessment (IQA) to assess observed reading comprehension instruction in the fall and spring (Matsumura et al., 2006; Matsumura, Garnier, Slater, & Boston, 2008). Observations were scheduled at a convenient time for teachers, and observers requested in

advance to visit the teachers' classrooms during the portion of their language arts block when the class held a discussion about a text. Observers engaged in a 3-day training session that consisted of studying the IQA rubrics along with videotaped lessons and practicing rating and taking field notes from longer excerpts of samples of videotapes. Observers who successfully completed that portion of the training then observed four to six classroom lessons with the data-collection supervisor. Observers who could not obtain over 80% agreement with the data-collection supervisor did not continue in the trial. Overall agreement between the observers and the data-collection supervisor across all observations for each rater pair was 86% in the fall and 87% in the spring.

Eight IQA scales measuring the quality of discussion participation and the discussion's academic rigor were used in the analyses. Each of the following ratings was assessed on a five-point scale (coded as 0 = poor to 4 = excellent): (*a*) students participating in the discussion, (*b*) teacher connects students' ideas to build coherence in the discussion, (*c*) students make contributions that link to and build on each other, (*d*) teacher presses students to explain their answers or support their assertions with evidence from a text, (*e*) students explain and support their answers, (*f*) teacher provides students with adequate time to fully express their thoughts, (*g*) academic rigor of text with the content and writer's craft providing enough complexity for discussion, and (*h*) academic rigor of the discussion with the teacher supporting the students to analyze and interpret a text. A single average was created from the individual ratings ($\alpha = .79$; see App. A for descriptions of lessons that received a 2 (fair) or a 3 (good) rating across these dimensions).

Texas Assessment of Knowledge and Skills (TAKS). Students' reading achievement was assessed by the Texas Assessment of Knowledge and Skills (TAKS), which was administered each spring in either Spanish or English depending on the individual child's language proficiency. Items on the TAKS assessment are grouped under four main objectives in which students: (1) demonstrate a basic understanding of culturally diverse written texts, (2) apply knowledge of literary elements to understand culturally diverse written texts, (3) use a variety of strategies to analyze culturally diverse written texts, and (4) apply critical thinking skills to analyze culturally diverse written texts.

Degrees of Reading Power Assessment (DRP). The DRP (Koslin, Zeno, & Koslin, 1987) is a group-administered test given to students in the fall and spring of each academic year that measures the most difficult text a student can read independently or with assistance. Within each passage, words are deleted and the student is asked to select the correct word for each deletion in the text. We chose the DRP because the assessment has a high level of technical quality and because the design of the test differs substantially from the TAKS.

Analyses. Repeated-measures ANOVAs were used to identify patterns of change from the fall of 2007 to the spring of 2008 in teachers' reports of the frequency of their participation in coaching activities, the content emphasized in the coaching activities, teachers' perception of the usefulness of coaching activities and content for improving their practice, and teachers' self-reports and observation ratings of the quality of reading comprehension instruction. One-way

ANOVAs were used to compare CFC and comparison teachers on measures within time.

In order to examine the influence of school participation in the CFC program on students' achievement, we analyzed a series of hierarchical linear growth models using HLM 6.0 (Raudenbush & Bryk, 2002). Although we explored the possibility of conducting a three-level analysis with these data, there were too few classrooms per school to reliably estimate variance between classrooms in a school.[2] Therefore, we present results from a two-level HLM model in which students were nested within schools. We chose this designation since it did not confound the classroom and school in the variance decomposition and since the CFC treatment was implemented at the school level.

Model building procedure. In the statistical models, we were primarily interested in the effect of CFC on students' spring achievement in 2008 (both the TAKS and the DRP) after controlling for their spring achievement measured in 2007 (TAKS and DRP, respectively). The random assignment of schools to treatment meant that there was a less than 5% chance that treated and comparison schools would not be equivalent on any measured or unmeasured covariates that might be related to the effect of the treatment on our outcome. However, in examining student-level covariates we found that there were significant differences between the treated and control groups prior to the study implementation. These covariates include the percent of fourth- and fifth-grade teachers participating in each school, the percent of ELL students in the school, and the percent of ethnic minority (Asian and African-American) students in the school. In order to adjust for these differences, we included them as covariates in the student level of our analyses.

To understand the influence of these covariates on our analyses, we employed a chi-square test to determine the trade-off between lost degrees of freedom and precision in our models. Thus, a model was determined to have better fit with the data if the chi-square test from successive models demonstrated a significant drop in deviance statistics relative to the degrees of freedom added to the model. Given this criterion, we retained the student-level covariates in our analyses. Additionally, preliminary analyses with the data had also indicated that there was reason to test for random effects across schools on the student-level covariates. We found that two covariates varied randomly between schools—the effect of being an ELL student significantly varied between schools ($p < .000$), as did the effect of prior achievement ($p = .05$). We retained the random effect for being an ELL student, but did not retain the random coefficient for the prior achievement effect since the model fit was better when this covariate was fixed (i.e., the chi-square test showed that the reduction in deviance did not merit the lost degrees of freedom).

Due to our preliminary analyses, we were interested not only in modeling the CFC effect on the intercept, but also on the effect of being an ELL student. We also examined whether or not there was an effect of CFC on any of the other student-level covariates, but no findings surfaced.[3] Therefore, we proceeded by omitting these parameters for model parsimony and also to retain a better model fit. The final model is a variation on the slopes-as-outcome model presented in Raudenbush and Bryk (2002) and is described below.

Student level:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{ELL}) + \beta_{2j}\,(\text{grade}) + \beta_{3j}\,(\text{Asian}) + \beta_{4j}\,(\text{African American}) + \beta_{5j}\,(\text{prior achievement}) + r_{ij}.$$

School level:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\,(\text{CFC}) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}\,(\text{CFC}) + u_{1j}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$
$$\beta_{5j} = \gamma_{50}.$$

In this model, Y_{ij}, the students' achievement score in the spring of 2008 (e.g., on the TAKS or DRP), is a function of the average achievement of students in school j (β_{0j}) plus the effects of student-level covariates (β_{1j} through β_{5j}), plus a random student-level error (r_{ij}) assumed to be normally distributed with a mean of zero and variance of σ^2. Of primary interest in these models is the effect of the CFC program, after adjusting for student-level covariates, on students' average scores (γ_{01}) and also on the effect of being an ELL student (γ_{11}). These coefficients describe whether the treatment resulted in greater achievement gains for (*a*) all students, (*b*) for the subset of ELL students, or (*c*) for both groups of students.

Results

What Is the Effect of a School's Participation in the CFC Program on Teachers' Experience of Coaching?

Frequency of participation in, and usefulness of, coaching. A significant ($p < .001$) main effect indicated that, averaged over both groups, teachers increased their overall participation in coaching activities over the year. Figure 1 shows that the increase in participation was significant for the CFC teachers, and that by spring, participation was significantly greater for teachers in the CFC schools than for teachers in the comparison schools. In spring, significantly more teachers in the CFC schools ($p < .01$) reported that participation in the coaching activities was useful for improving their practice ($\text{mean}_{\text{CFC}} = 2.87$) compared to the comparison teachers ($\text{mean}_{\text{comparison}} = 1.98$).

Looking within the individual coaching activities, CFC teachers reported significantly greater participation in all five coaching activities: coaches met with them and other teachers in grade-level meetings ($p < .001$), coaches met with them individually ($p < .001$), coaches observed them teaching for at least 30 minutes ($p < .001$), coaches taught a model lesson for them ($p < .001$), and

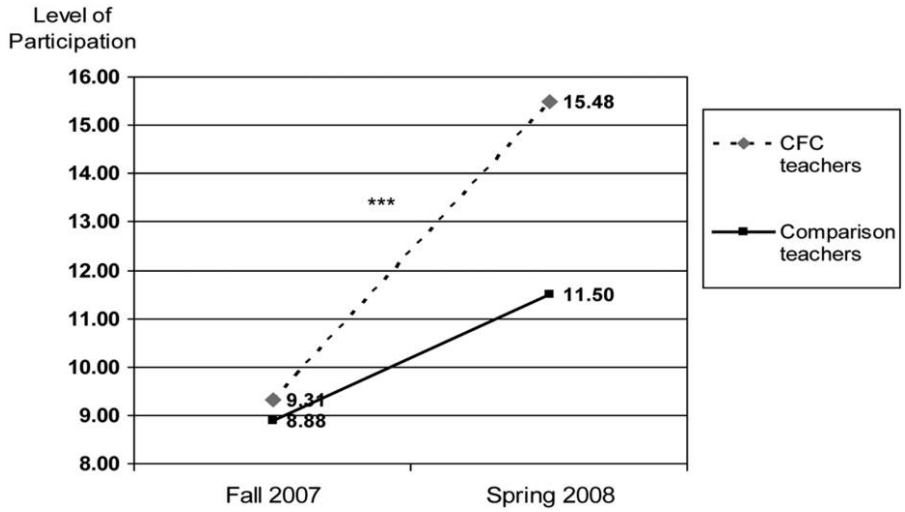

Figure 1. Overall teacher participation in literacy coaching activities, fall 2007 to spring 2008. Significant change over time indicated in graph by $***p < .001$. Within-time comparisons indicated the following difference: fall 2007 ($SD_{CFC} = 5.99$; $SD_{comparison} = 5.51$): no difference detected; spring 2008 ($SD_{CFC} = 3.58$; $SD_{comparison} = 3.90$): CFC > comparison, $p < .01$.

coaches taught a lesson with them in their classroom ($p < .01$). Although the CFC and comparison teachers did not differ at baseline, by spring the CFC teachers reported that their coach more frequently met with them in grade-level meetings ($p < .05$), observed them teaching for at least 30 minutes ($p < .01$), and taught a model lesson ($p < .01$). Table 1 describes the percent of teachers in the CFC and comparison groups participating in individual coaching activities at different levels.

While Cohort 2 teachers in the treatment schools participated in coaching more frequently than their counterparts in the comparison schools, relatively few teachers participated in coaching at the level intended by the developers of the CFC program. As described earlier, CFC coaches in this study were expected to meet with teachers in weekly grade-level team meetings or individually to study the theory underlying effective reading comprehension instruction and to meet with teachers individually on a monthly basis at minimum to engage in a cycle of lesson planning, enacting, and/or reflecting on instruction. About half of the Cohort 2 teachers in the CFC schools (48%) participated in coaching at or near this intended level, compared with 23% of the teachers in the comparison sample.

Content emphasized in the coaching activities. Analyses identified significant ($p < .001$) time main effects in all content areas emphasized in the coaching activities. Table 2 shows that by spring teachers in CFC schools described significantly greater emphasis than teachers in comparison schools in the four content areas targeted by the CFC program: planning and reflecting on instruction ($p < .01$), providing help during lesson enactments ($p < .05$), building knowledge of the theory underlying effective reading comprehension instruction ($p < .05$), and differentiating instruction relative to their peers in the comparison schools ($p < .01$). Teachers in the CFC schools reported significant ($p < .001$) increases in all of these areas.

Table 1. Percent of CFC and Comparison Teachers Participating at Different Frequencies in Literacy Coaching Activities, Fall 2007 to Spring 2008 (Cohort 2 Teachers; $N = 73$)

	CFC Group					Comparison Group				
Literacy Coaching Activity	Never	1–3 Times per Year	4–6 Times per Year	Monthly	Weekly	Never	1–3 Times per Year	4–6 Times per Year	Monthly	Weekly
Fall 2007:										
Coach met with me and other teachers in grade-level meetings	71	0	7	10	13	55	18	5	18	5
Coach met with me individually	77	10	3	3	7	77	9	0	5	9
Coach observed me teaching for at least 30 minutes	74	13	0	7	7	86	5	5	0	5
Coach taught a model lesson for me	87	7	3	0	3	82	14	0	5	0
Coach taught a lesson with me in my classroom	90	3	0	3	3	82	14	0	5	0
Spring: 2008										
Coach met with me and other teachers in grade-level meetings	0	7	10	28	55	0	18	23	36	23
Coach met with me individually	2	22	28	28	20	5	36	23	23	14
Coach observed me teaching for at least 30 minutes	10	39	20	26	5	55	18	14	9	5
Coach taught a model lesson for me	8	46	23	13	10	59	27	9	0	5
Coach taught a lesson with me in my classroom	30	43	15	10	2	68	9	14	0	9

Table 2. Results of Repeated-Measures Analyses of Teacher Reports of Emphasis of Literacy Coaching Activities, Fall 2007 to Spring 2008 (Cohort 2 Teachers; $N = 73$)

Emphasis of Literacy Coaching Activity	Fall 2007		Spring 2008		Repeated-Measures ANOVA		
	CFC	Comparison	CFC	Comparison			
	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)	Group	Time	GXT
Planning and reflecting on instruction	1.50 (.82)	1.25 (.49)	2.77[a] (.86)	1.92[a] (.71)	**	***	*ns*
Help during lesson enactment	1.50 (.87)	1.29 (.49)	2.91[b] (.83)	2.24[b] (.79)	*	***	*ns*
Building theory underlying pedagogy	1.43 (.86)	1.27 (.80)	2.40[b] (1.07)	1.67[b] (.98)	*ns*	***	*ns*
Differentiating instruction	1.53 (.94)	1.23 (.56)	2.45[a] (.91)	1.60[a] (.70)	*ns*	**	*ns*

Note.—Teachers rated each item from 1 = no emphasis to 4 = a great deal of emphasis.
[a] CFC > comparison, $p < .01$.
[b] CFC > comparison, $p < .05$.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Do Teachers in Schools Participating in the CFC Program Improve the Quality of Their Reading Comprehension Instruction?

Teachers' self-reported practice. Analyses identified significant ($p < .001$) time main effects for average measures of teachers' self-reported reading comprehension instruction. From the fall of 2007 to the spring of 2008, teachers in both CFC and comparison schools described their students as communicating more frequently with each other in class discussions ($p < .01$ and $p < .05$, respectively), but only teachers in CFC schools described students more often connecting their responses to ideas expressed by other students during discussions ($p < .01$). By spring, CFC teachers described more frequent communication among their students than did comparison teachers ($p < .05$); no difference was detected between groups in the extent to which students made connections with ideas in class discussions.[4]

To gain a deeper understanding of the different ways in which teachers' instructional practices changed, we turned to teachers' reports of specific instructional practices in their classrooms related to class discussions about a text. Table 3 describes these practices and the percents of teachers who used these practices in their lessons often or almost always, rarely or sometimes, or never. By spring, CFC teachers reported significantly more occurrences of six student activities related to higher-quality reading comprehension instruction. Over 20% of CFC teachers who rarely or sometimes reported these activities now reported they often or almost always occurred in their classrooms. Four of these activities involved students applying higher-level reasoning skills to comprehend a text: (*a*) connecting ideas and events within a text, (*b*) making connections across texts, (*c*) using examples from a text to support their ideas, and (*d*) identifying the author's meaning. Two other activities involved students communicating with each other dur-

Table 3. Percent of Cohort 2 CFC Teachers Reporting Frequency of Student Activities during Class Discussions, Fall 2007 and Spring 2008 (Cohort 2 Teachers; $N = 73$)

	Fall 2007			Spring 2008		
Student Activity[a]	Never	Rarely/ Sometimes	Often/ Almost Always	Never	Rarely/ Sometimes	Often/ Almost Always
Students make connections with ideas:						
Students connect ideas and events within a text[b]	4	46	50	0	11	89
Students make connections across texts[c]	4	43	54	0	11	89
Students use examples from a text to support their ideas[c]	4	41	56	0	14	86
Students identify the author's meaning[b]	7	50	43	0	36	64
Students draw on relevant academic knowledge to support their assertions (e.g., social studies, science, etc.)	4	50	46	0	36	64
Students get off the subject being discussed in class[c]	4	86	11	0	86	14
Teacher and/or students summarize the main points of a discussion	4	15	82	0	18	82
Students communicating with their peers:						
Students build on each other's ideas during discussion	4	32	64	0	11	89
Students provide constructive feedback to their peers[c]	11	44	44	0	32	68
Most students participate in the discussion at some point[b]	4	32	64	0	18	82
Students show each other respect during a discussion	4	14	82	0	4	96

[a] Codes: 1 = never, 2 = rarely, 3 = sometimes, 4 = often, 5 = almost always.
[b] Significant mean change from fall 2007 to spring 2008, $p < .01$.
[c] Significant mean change from fall 2007 to spring 2008, $p < .05$.

ing class discussions: (*a*) students provide constructive feedback to their peers, and (*b*) most students participate in the discussion.

Eighteen to twenty-four percent more CFC teachers reported that by spring these activities occurred often or almost always. We also found that 3% more CFC teachers reported that their students often or almost always got off topic in class discussions. It is possible that including more of the students in class discussions and asking students more open-ended questions (e.g., with regard to how a text relates to another text)—both goals of the CFC program—may have increased the likelihood that some students would stray off topic.

Comparison teachers reported significantly ($p < .05$) more occurrences of just one activity. By spring, 6% more comparison teachers reported that students in their classes connected ideas and events with a text often or almost always.

Observed reading comprehension instruction. Observers assessed the quality of classroom text discussions in both the fall and spring. The average rating over time for both groups of teachers consistently fell between fair and good, ranging

Table 4. Results of HLM Analyses

	TAKS (N = 1,229 students)		DRP (N = 896 students)	
Final Fixed Effects	Coefficient	*SE*	Coefficient	*SE*
Intercept, γ_{00}	2,157.96***	7.05	42.02***	.74
CFC program, γ_{01}	21.02	13.75	−.02	1.50
ELL student, γ_{10}	−35.71**	11.32	−2.23*	.82
CFC program, γ_{11}	59.84*	22.33	.58	1.71
Grade level, γ_{20}	68.16***	8.52	4.67**	1.65
Asian, γ_{30}	26.21	20.88	4.52***	1.12
African American, γ_{40}	−48.53***	10.89	−1.74	1.08
Prior reading achievement, γ_{50}	.69***	.03	.66***	.04

* $p < .05$.
** $p < .01$.
*** $p < .001$.

from 2.28 to 2.64 (examples of lessons with fair and good quality of text discussions are presented in App. A). A significant ($p < .01$) group main effect indicated that, averaged over time, the quality of instruction in the CFC schools was rated higher than in the comparison schools. Unexpectedly, the quality of instruction was observed at a significantly higher level for CFC teachers than comparison teachers in both the fall and spring ($p < .05$ each time), and no change in instruction was detected for either group. Scheduling the fall observations after the CFC program was underway in trial year 2 may partially explain this finding and is considered in the discussion section.

Do Students in Schools Participating in the CFC Program Improve Their Reading Comprehension Skills?

In order to understand the influence of CFC on student learning, we began by examining students' 2008 scores on the TAKS while adjusting for their prior achievement. These models also included other student-level covariates. These results, displayed in the left-hand column of Table 4, show that, on average, students in the fifth grade scored higher than students in the fourth grade and African American students scored lower than other students. Additionally, ELL students, representing about 40% of the students in our sample, also scored lower than non-ELL students, on average. The covariates explained 49% of the variance in the TAKS.

Our primary interest in these models was the effect of the CFC treatment on the learning of all students in general (γ_{01}), and on the learning of ELL students in particular (γ_{11}). Although the coefficient for the CFC program on the TAKS intercept was positive, there was no significant effect ($p = .14$). The results, however, do indicate a significant effect of the CFC program on the ELL effect ($p = .01$). While ELL students, on average, scored 40 points lower than native English-speaking students (γ_{10}), ELL students in CFC schools scored almost 60 points higher than ELL students in comparison schools. Whereas ELL students scored dramatically lower than native English speakers in comparison schools, ELL students in CFC schools scored almost on par with their non-ELL peers.

Table 5. Percent of Variance Explained by Addition of CFC as a Predictor in Spring 2008 TAKS and DRP Achievement Scores

	TAKS				DRP			
Random Effects	Variance Components	χ^2	p	% Variance Explained	Variance Components	χ^2	p	% Variance Explained
Initial random effects:								
Intercept, u_{oj}	1,097.75	81.45	.000		11.70	143.43	.000	
Slope for ELL students, u_{1j}	2,126.08	52.48	.000		5.53	31.28	.051	
Sigma squared, r_{ij}	14,787.85				60.09			
Final random effects:								
Intercept, u_{oj}	1,013.76	73.86	.000	8	12.26	143.42	.000	0
Slope for ELL students, u_{1j}	1,690.40	47.28	.001	21	6.33	31.49	.035	0
Sigma squared, r_{ij}	14,765.01				60.09			

This effect is noteworthy because of the large percentage of ELL students in the district as well as the size of the effect. On average, students' scale scores on the TAKS had a mean of 2,121.60 and a standard deviation of 117.32. This means that the coefficient for CFC on the ELL slope of 59.84 translates into an effect size of .51 for the CFC program on ELL students. Finally, Table 5 shows the additional variance explained for being a CFC school after adjusting for the student-level covariates, including prior achievement. The addition of CFC to the model explained 8% of the remaining between-school variance in students' achievement scores and 21% of the between-school variance that the effect of being an ELL student had on achievement.

Similar models were examined for the effects of the CFC program on students' DRP scores. Only 896 students had data available for these analyses. The results obtained using the DRP as an outcome (shown in the right-hand column of Table 4) provide different results from the TAKS. While there remains a significant effect for being in the fifth grade as well as a significant negative effect of being an ELL student, on average, some differences also exist between the models. After adjusting for the other covariates in the model, African American students score no different than white students, and Asian students score higher than other students. The covariates explained 48% of the variance in the DRP. Using students' DRP scores as an outcome, there is no effect of the CFC program on either the intercept or ELL effect. The addition of CFC did not explain an additional proportion of variance.

Summary

Overall, our results indicate that the CFC program engaged the newly recruited (Cohort 2) teachers in key literacy coaching activities over the school year that

translated into effective instructional practices in the classroom. In schools participating in the CFC program, the quality of teachers' reported and observed instructional practices for classroom discussions improved beyond those in comparisons schools, and, on average, these schools also achieved significant learning gains for their ELL students.

Specifically, we found that teachers in the CFC schools significantly increased the frequency of their participation in coaching relative to the teachers in the comparison schools, and were more likely to view their participation in coaching as useful for improving their instructional practice. Commensurate with the goals of the CFC program, CFC coaches emphasized planning and reflecting on instruction, providing help during lesson enactment, understanding the theory underlying effective reading comprehension instruction, and differentiating instruction in their work with teachers more than coaches in the comparison schools.

Despite the overall increase in teachers' participation in activities focused directly on building their knowledge base and pedagogical skills, only half of the teachers in the CFC schools participated in coaching at or near the level intended by the program developers.[5] One explanation for this pattern may be that participation in coaching was voluntary in both the treatment and comparison schools. Principals could strongly encourage teachers to meet with their coach but, at the end of the day, teachers generally made the final decisions about the frequency and depth of their interactions with the coaches. Moreover, even in the CFC schools, coaching activities were an addition to teachers' regular duties and therefore competed for time with the teachers' other professional obligations. Unsurprisingly, lack of time has topped the list of reasons teachers identify for participating in coaching less frequently (Matsumura, Sartoris, Bickel, & Garnier, 2009). Other reasons have included teachers' belief that they would not benefit from coaching because their practice is fine as it is and their belief that the CFC program is just another program in a long list of reforms that eventually would pass on (Matsumura, Sartoris, et al., 2009).

Even with the lower-than-planned level of participation in coaching, we found significant improvement in the quality of teachers' self-reported classroom text discussions over the year in the CFC schools. Teachers in these schools described students as more frequently discussing texts with each other and exhibiting higher-level conceptual understanding by making more connections with ideas and events within and across texts. The observed instructional quality in CFC schools, however, did not change since observers rated CFC schools significantly higher than comparison schools in both the fall and spring.

Finally, at the end of the year, we found significant average gains on the state standardized test for the CFC schools. In contrast to the comparison schools, where the ELL students scored far lower than their non-ELL counterparts, the ELL students in the CFC schools scored higher than the ELL students in the comparison schools and almost on par with their non-ELL peers. We did not find a significant effect for all students (ELL and non-ELL), though the trend in the data was in a positive direction for the CFC schools. No difference between the comparison and CFC schools was detected on the DRP assessment.

Discussion

Teacher mobility poses a major challenge to implementing comprehensive PD programs (such as CFC) that seek to increase student achievement by engaging teachers in sustained efforts to improve their practice. Faced with the unexpected loss of half the sample of teachers at the end of the first year, we recruited a replacement sample of teachers in the treatment and comparison schools. Following cohorts of teachers with more and less exposure to CFC, and at different stages of implementation, provided an opportunity to develop a more nuanced understanding of the potential for effective implementation of the program in schools with high teacher mobility. The relative effects of the program on the two teacher cohorts and their students are discussed in the following sections. The implications of our findings for optimally implementing literacy coaching programs as well as using randomized controlled designs to study their effectiveness in schools with high teacher mobility are also explored.

Students' Reading Achievement and Coaching

An important finding of this study is the positive, significant effect of the CFC program on the school-level reading achievement gains of ELL students after just 1 year of teacher participation. This finding mirrors the effect we found for the ELL students of the Cohort 1 teachers in the CFC schools after 2 years of their participation. We returned to the student reading achievement scores of Cohort 1 teachers to answer the question of whether their participation at the start of the CFC program produced a similar effect on their students' learning by the end of their first year of the program. Using identical procedures for analyzing models of student achievement that we used for Cohort 2, we found that Cohort 1 teachers' participation in the first year of the CFC program was not associated with gains in student learning at the school level.

One explanation for the school gains in ELL students' reading achievement that we found for both teacher cohorts within the CFC schools may be that the ELL students who are achieving at a lower level than the full sample of students may have more room to grow. Another explanation is that ELL students may benefit particularly from the instructional practices promoted in the CFC program, notably, increased opportunities for productive language use in class discussions (Goldenberg, 2008; Shanahan & Beck, 2006). Evidence from the teacher surveys and ratings of observed text discussions indicates that a greater percentage of classrooms in the CFC schools received good overall ratings of classroom discussions compared to the comparison schools (see App. A for a description of the differences between good and fair ratings of class discussions). Aligned with the instructional goals of Questioning the Author, more students in the CFC schools participated in class discussions characterized by teachers encouraging student-to-student interactions, pressing students to provide evidence to support their assertions, and posing more open-ended interpretive questions. Students involved in these discussions were more likely to build on one another's responses, explain their answers, and use evidence from a text to support their assertions in

class discussions in the CFC schools than those students in the comparison schools.

Teachers' Experiences of Coaching

To understand why the effects of the CFC program on learning appeared more quickly for the newly recruited Cohort 2 teachers, we turn to teachers' experiences in the program. As with the Cohort 1 CFC teachers, the Cohort 2 CFC teachers significantly increased their participation in coaching activities focused directly on building the teachers' knowledge base and improving their instruction during their first year, and they felt more positive about the usefulness of coaching for improving their teaching relative to teachers in the comparison schools. Unlike the Cohort 1 teachers, the Cohort 2 teachers in the CFC schools benefited from an earlier start working with their coaches at the beginning of the academic year. In the first year of the trial, CFC coaches were not introduced to their school until November, well after the school year started and close to the winter holiday and state accountability testing that began in February. Such a late start severely limited the time available for teachers to work with coaches, especially considering state accountability testing in spring that reportedly demands a major portion of the teachers' time. The effect of the CFC program on student learning in schools likely appeared in half the time for the Cohort 2 teacher sample because these teachers entered the CFC schools with an already well-established coaching program and experienced coaching activities much earlier in their first year.

The Cohort 2 CFC teachers not only began their first year in the schools with an established coaching program, but interacted daily with a principal and colleagues familiar with the CFC program and Questioning the Author. A main goal of the CFC program is building a professional learning community within the schools. In many of the CFC schools, the newly recruited Cohort 2 teachers experienced a different professional culture in their first year than did the Cohort 1 teachers. Interviews with coaches indicated that most of them established a more regular pattern of meeting with teachers in collaborative settings in the second year of the trial, compared to the first trial year in which most of their efforts focused on building relationships with principals and teachers (Matsumura, Sartoris, et al., 2009). Further analyses rendered evidence that the norms for professional collaboration changed in the CFC schools after the first year of the program. Compared to the Cohort 1 teachers in the CFC schools in their first year, significantly more Cohort 2 teachers in the CFC schools described teachers in their school regularly discussing teaching and learning ($p < .05$), discussing instruction with colleagues ($p < .05$), inviting each other into their rooms to observe and give feedback ($p < .05$), sharing instructional material and lesson plans ($p < .05$), and having a can-do attitude ($p < .01$). Quite possibly, there is an added benefit of an established, comprehensive coaching program that aims to develop a sustainable learning community within a school. Such an established program has the potential for bringing large numbers of incoming teachers, who tend to be less experienced, into the school community and more quickly up to speed on their school's instructional reforms.

Quality of Reading Comprehension Instruction and Coaching

Both teacher cohorts in the CFC schools exhibited higher-quality reading instruction in their classrooms by the end of the second year of the program. Even though we found a link between the CFC program and improved quality of instruction for Cohort 1 teachers in the CFC schools (Matsumura, Garnier, et al., 2009), the evidence is not as clear for the second cohort. The Cohort 2 CFC teachers exhibited a higher level of instruction in their classrooms than comparison teachers at baseline and at the end of the study year.

Why did raters see a higher quality of CFC teachers' reading instruction in Cohort 2 than in Cohort 1 classrooms near the beginning of the academic year when they first entered the program? Perhaps they were more skilled at teaching reading comprehension. This explanation is possible but unlikely given the similarities between the two cohorts on all other measures at baseline (e.g., education, years of experience teaching, certification, past participation in literacy coaching, and self-reported quality of class discussions). By spring, both cohorts of teachers in the CFC schools described significant improvement in the quality of their classroom discussions, with their students participating more in public discussions, communicating more often with their classmates, and making more conceptual links.

A more likely explanation can be located in the timing of the baseline classroom observations. The first observations of Cohort 2 classrooms occurred in October after 2 months of participation in the CFC program and therefore cannot be considered true baseline observations. As discussed earlier, the Cohort 2 CFC teachers, unlike the Cohort 1 CFC teachers, entered schools with an already established coaching program. By October, the Cohort 2 teachers in the CFC schools already had attended a 3-day PD in-service led by the coaches, which was intended to familiarize them with Questioning the Author, and had participated in 6–8 weeks of coaching. These experiences, in tandem with interacting daily with teacher colleagues who were already familiar with the CFC program and Questioning the Author, likely generated a faster integration of the information that the new teachers learned from their coaches. Combined with delays observing classrooms in the fall, the same processes that led to a quicker uptake of the instructional processes promoted by CFC and school improvement in the reading achievement of ELL students probably undermined our ability to obtain a true baseline measure of the Cohort 2 teachers' practice.

Implications for Implementing Literacy-Coaching Programs

We originally based the longitudinal design of the CFC program intervention on the expectation that 3 years would be needed for positive effects on instruction and learning to surface. Numerous studies in the social sciences indicate that at least this length of time is needed for complex multilevel interventions to influence desired outcomes (studies summarized in Fixsen, Naoom, Blasé, Friedman, & Wallace, 2005). However, other research indicates that less time is needed, suggesting that only 30 hours of PD may be needed to bring about instructional improvement (Guskey & Yoon, 2009). Given the demands on teachers' time in

many schools around end-of-the-year accountability testing, effects for teachers who are new to a school may be maximized only if a coaching program is well established and engages teachers in intensive PD at the very beginning of the school year. We suggest that modest improvements in instruction and learning can occur in a shorter time span than originally hypothesized—depending on the nature of the learning environment and stage of program implementation.

Our findings also suggest that coaches would benefit from programs that provide them with support for working continuously with shifting populations of teachers and integrating new teachers into existing collaborative structures without losing focus on continuing teachers. Questions still remain about the extent to which coaching can be expected to improve instruction and learning under conditions of high teacher mobility. Research described earlier in this article indicates that high levels of teacher mobility can have a deleterious effect on school-level instructional reform. Under conditions of high teacher turnover, school leaders' attention can get diverted from deepening and sustaining existing reforms toward hiring and mentoring teachers who are new to the school (Allensworth et al., 2009). A major concern in implementing coaching programs is helping coaches who are faced with large numbers of incoming teachers every year and may spend the bulk of their time helping these teachers catch up. This demand could diminish their ability to further the ongoing development of teachers who worked with them the previous year.

Finally, it is worth considering that the goal of building a learning community as an essential part of a coaching program could have multiple benefits. Coaches and the existing community of colleagues can work together to bring new teachers more quickly into the program. Such a goal may contribute beyond the immediate aims of improving instruction and learning to increasing teacher retention. Research indicates that teachers are more likely to remain in hard-to-staff schools when they are part of a collaborative culture (Allensworth et al., 2009). If teachers feel they are learning instructional practices that will help them professionally and increase their confidence and ability to cope with at-risk students, then teachers who would be more likely to leave may decide to remain in their schools. Retaining teachers ultimately reduces the costs for schools of hiring and training replacement teachers.

Implications for Conducting Randomized Field Trials

Conducting rigorous randomized trials of complex interventions in natural field conditions presents substantial challenges to the researcher. Our randomized trial exhibited high teacher turnover—a key link between the literacy-coach training intervention (CFC) and study outcomes. To preserve statistical power for cross-sectional analyses and avoid selection bias issues, we recruited new teachers to replace those—in both treatment and comparison schools—who left our trial due to natural turnover. An advantage of recruiting teachers midway through the program's implementation is that it creates an opportunity to investigate whether the effects of a program detected with one cohort of teachers can, in a sense, be replicated with a second cohort. Analyzing the multiple teacher cohorts together as a "staggered start" design (Loughlin, 2006) also allows one to separate longitu-

dinal effects of a reform program such as CFC from the effects of transient changes in the school district environment (e.g., the development of competing PD programs in a district, change in principal leadership, etc.).

A challenge of these analyses is that one of the desired outcomes of many comprehensive reform programs—to transform the professional community in schools—can make obtaining clean baseline data on subsequent teacher cohorts' practice very difficult. The very nature of a comprehensive reform program creates the conditions for diffusion of treatment threats to a longitudinal study's design when data are collected on multiple cohorts. An obvious solution to this problem is to collect baseline data very early in the academic year. This can be very difficult to accomplish, however, when multiple data are being collected within the same time period. Losing large numbers of teachers across years also greatly reduces the number of teachers who can be studied over 3 years or longer, thus weakening a study's ability to detect effects over time and/or investigate the sustainability of a reform's impact on practice. Recruiting significantly greater numbers of schools is a potential solution, although this can greatly increase the cost of the research and limit the locations where such research could be conducted to very large districts.

Conclusion

Increasingly higher rates of teacher mobility present an enormous challenge and cost to schools attempting to improve instruction through sustained and ongoing PD for teachers. Our evidence confirms that an established comprehensive coaching program can improve a school's reading achievement for ELL students whose teachers are new to their school in part because an established program is in place and ready to train teachers. Research is needed to assess the degree to which instruction and learning can continue to improve in schools in which a quarter to half of the teaching staff is new each year. Additional research is also needed that systematically investigates the effect of high teacher mobility on coach-teacher interactions. In schools with high teacher turnover, do coaches continue to work intensively with the teachers they have worked with in the past, or do they direct more attention toward the teachers they perceive as needing support (i.e., novice teachers and teachers who are new to the school)?

A comprehensive coaching program, such as CFC, may further benefit schools by indirectly reducing teacher mobility. In our study, similar levels of mobility appeared across the conditions, suggesting that CFC did not reduce teacher mobility by the end of the first study year. Research indicates, however, that some of the CFC program goals (e.g., fostering strong collaborative relationships among teachers and between principals and teachers) overlap with the conditions associated with teacher retention in schools serving high numbers of poor and minority students (Allensworth et al., 2009). Conceivably, a coaching program such as CFC could help reduce attrition in schools over time if the program was successful in achieving these key goals. Moreover, given the difficulty of creating coherent and sustained professional learning opportunities in schools with large, ever-shifting teacher populations, reducing teacher attrition could be an important

proximal outcome to consider when assessing the effectiveness of coaching programs.

Appendix A

Description of Text Discussions with Variable Ratings

Fair Quality Text Discussions

Text discussions considered to be of fair quality (i.e., received twos across most dimensions) are characterized by teachers asking closed-ended queries that guide students to build a coherent, albeit surface-level understanding of the events in a text. Students are not asked questions that guide them to infer meaning, make connections across texts, or link the text to their own experiences. Students provide brief answers (often only a few words) to questions posed by the teacher. Students do not build on each other's contributions or take an active role in shaping the direction of the conversation.

Good Quality Text Discussions

In class discussions that receive good (not excellent) overall ratings (i.e., threes across most dimensions) there is some evidence that teachers use techniques that are intended to foster authentic discussions in classrooms. At least a few times in the discussion teachers actively encourage students to respond to each other's statements by connecting students' comments (e.g., "I hear some disagreement. Chris is saying yes. David is saying no") and by marking students' contributions (e.g., "Oh, Brianna is making a connection. The dad named her India because he had been there like she named the dog after the store"). At least a few times in the discussion, teachers pose an interpretive question that encourages multiple answers from students (e.g., "Why do you think [the author] says, 'In my mind he's the preacher'?") and press students to explain their answers (e.g., "Is that an important part? Why?").

Notes

The work described in the current article was supported by a Teacher Quality Grant from the Institute for Educational Sciences, R305W060027. We are appreciative of the support provided by IES. All errors of fact, omission, and/or interpretation are solely the responsibility of the authors.

1. We involved principals and district staff in the assignment of schools to conditions so that participants would know that the selection process was unbiased. The district convened a meeting of the principals of the 32 eligible schools. After learning about the CFC program and the study from district leaders and the principal investigator, interested principals ($N = 30$) completed consent forms, and the name of their school was placed in a container. A principal from the audience drew names from the container, alternating with each selection whether the school would participate in the CFC program or serve as a control sample.

2. It is important to note that the parameter estimates obtained from either the three-level model or the alternate two-level model with students nested in classrooms are nearly identical to those presented here. In other words, the decision to present one type of model over another did not in any way influence the primary inference made from our analyses.

3. Nor did the inclusion of CFC on any of the student-level covariates in any way alter our primary findings.

4. ANOVAs indicated no differences in teacher education or experience teaching, reports of previous PD activities, reported instructional practices, observed quality of instruction, or student reading achievement between the teachers in the comparison schools who did or did not have coaches. For this reason, all of the teachers in the comparison schools were included in the analyses reported in this and the following sections.

5. Approximately a quarter of the teachers in the comparison schools participated in coaching at the level specified by the CFC program. The district did not communicate a specific standard for how frequently literacy coaches were to meet with teachers in comparison schools.

References

Allensworth, E., Ponisciak, S., & Mazzeo, M. (2009). *The schools teachers leave: Teacher mobility in Chicago public schools.* Chicago: Consortium on Chicago School Research.

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth.* Mahwah, NJ: Erlbaum.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey-Bass.

Beck, I. L., & McKeown, M. B. (2006). *Improving comprehension with Questioning the Author: A fresh and expanded view of a powerful approach.* New York: Scholastic.

Beck, I. L., McKeown, M. G., Hamilton, R. L., & Kucan, L. (1997). *Questioning the Author: An approach for enhancing student engagement with text.* Newark, DE: International Reading Association.

Biancarosa, G., Bryk, A., & Dexter, E. (2008, April). *Assessing the value-added effects of literacy collaborative professional development on student learning.* Paper presented at the annual meeting of the American Educational Research Association, New York.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). *Who leaves? Teacher attrition and student achievement* (NBER Working Paper No. 14022). Retrieved October 22, 2009, from http://www.nber.org/papers/w14022

Camburn, E. M., Kimball, S. M., & Lowenhaupt, R. (2008). Going to scale with teacher leadership: Lessons learned from a district-wide literacy coach initiative. In M. M. Mangin & S. R. Stoelinga (Eds.), *Effective teacher leadership: Using research to inform and reform* (pp. 120). New York: Teachers College Press.

Chall, J., & Jacobs, V. A. (2003). Poor children's fourth-grade slump. *American Educator,* **27**(1), 14–15, 44.

Collins, A., Brown, J. S., & Holum, A. (1991). Cognitive apprenticeship: Making thinking visible. *American Educator,* **6**(11), 38–46.

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

Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading First impact study final report* (NCEE 2009-4038). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement* (NCEE Report 2008-4031). Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences.

Goldenberg, C. (2008). Teaching English Language Learners: What the research does and does not say. *American Educator,* **32**(2), 8–23, 42–44.

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

Henusheck, E. A., Kain, J. F., & Rivkin, S. (2004). Why public schools lose teachers. *Journal of Human Resources,* **39**(2), 326–354.

Joyce, B., & Showers, B. (1996). The evolution of peer coaching. *Educational Leadership,* **53**(6), 12–16.

Knight, J. (2004). Instructional coaches make progress through partnership. *Journal of Staff Development,* **25**(2), 32–37.

Koslin, B. L., Zeno, S., & Koslin, S. (1987). *The DRP: An effectiveness measure in reading.* Brewster, NY: Touchstone Applied Science Associates.

Loughlin, T. M. (2006). Improved experimental design and analysis for long-term experiments. *Crop Science,* **46**, 2492–2502.

Marsh, J. A., McCombs, J. S., Lockwood, J. R., Martorell, F., Gershwin, D., Naftel, S., et al. (2008). *Supporting literacy across the sunshine state: A study of Florida middle school reading coaches.* Santa Monica, CA: RAND.

Marvel, J., Lyter, D. M., Peltola, P., Strizek, G. A., & Morton, B. A. (2006). *Teacher attrition and mobility: Results from the 2004–05 Teacher Follow-Up Survey* (NCES 2007-307). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Matsumura, L. C., Garnier, H. E., Junker, B., Resnick, L., & Bickel, D. B. (2009, March). *The influence of Content-Focused Coaching on reading comprehension instruction and student achievement.* Paper presented at the annual meeting of the Society for Research on Educational Effectiveness, Washington, DC.

Matsumura, L. C., Garnier, H., Slater, S. C., & Boston, M. B. (2008). Measuring instructional interactions "at-scale." *Educational Assessment,* **13**(4), 267–300.

Matsumura, L. C., Sartoris, M., Bickel, D. B., & Garnier, H. E. (2009). Leadership for literacy coaching: The principal's role in launching a new coaching program. *Educational Administration Quarterly,* **45**(5), 655–694.

Matsumura, L. C., Slater, S. C., Junker, B., Peterson, M., Boston, M., Steele, M., et al. (2006). *Measuring reading comprehension and mathematics instruction in urban middle schools: A pilot study of the Instructional Quality Assessment* (CSE Technical Report No. 681). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

McLaughlin, M. W., & Talbert, J. (2006). *Building school-based teacher learning communities: Professional strategies to improve student achievement* (Series on School Reform). Chicago: University of Chicago Press.

Neufeld, B., & Roper, D. (2003). *Coaching: A strategy for developing instructional capacity. Promises and practicalities.* Washington, DC: Aspen Institute Program on Education and Annenberg Institute for School Reform.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Reichardt, S. (2008). *Teacher attrition and teacher diversity in Colorado: State of teaching in Colorado.* Denver: Center for Education Policy Analysis, University of Colorado Denver.

Sailors, M., & Price, L. (2010). Professional development that supports the teaching of cognitive reading strategy instruction. *Elementary School Journal,* **110**(3), 301–322.

Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2005). *Do teachers really leave for higher paying jobs in alternative occupations?* (Economics CIBC Human Capital and Productivity Project Working Paper Series, Paper 11). Retrieved October 22, 2009, from http://ir.lib.uwo.ca/economicscibc/11

Shanahan, T., & Beck, I. L. (2006). Effective literacy teaching for English language learners. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth* (pp. 415–489). Mahwah, NJ: Erlbaum.

Snow, C. (2002). *Reading for understanding: Toward a R&D program in reading comprehension.* Santa Monica, CA: Rand Education.

Staub, F. C., & Bickel, D. D. (2003, August). *Developing Content-Focused Coaching in elementary literacy: A case study on designing for scale.* Paper presented at the biannual meeting of the European Association on Research and Instruction (EARLI), Padova, Italy.

Tharp, R. G., & Gallimore, R. (1988). *Rousing minds to life: Teaching, learning and schooling in social context.* New York: Cambridge University Press.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad.* Dallas: National Staff Development Council.

West, L., & Staub, F. C. (2003). *Content-Focused Coaching: Transforming mathematics lessons.* Portsmouth, NH: Heinemann.