Learning and Instruction 25 (2013) 35–48

Contents lists available at SciVerse ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Literacy coaching to improve student reading achievement: A multi-level mediation model

Lindsay Clare Matsumura [a,*], Helen E. Garnier [b,1], Jessaca Spybrook [c,2]

[a] Learning Sciences and Policy, University of Pittsburgh, 5808 WWPH, Pittsburgh, PA 15260, USA
[b] BSCS/Learning Research and Development Center, 5808 WWPH, Pittsburgh, PA 15260, USA
[c] Western Michigan University, 1406 Sangren Hall, Kalamazoo, MI 49008, USA

ARTICLE INFO

Article history:
Received 13 January 2012
Received in revised form
24 October 2012
Accepted 5 November 2012

Keywords:
Literacy coaching
Mediation analysis
Literacy instruction
Reading achievement

ABSTRACT

In a longitudinal group-randomized trial, we explore the key role of the quality of classroom text discussions in mediating the effects of Content-Focused Coaching (CFC) on student reading achievement (2983 students, 167 teachers). Schools in the United States serving large numbers of minority and English language learning (ELL) students from low-income families were randomly assigned to participate in the CFC literacy-coaching program or to continue with the literacy coaching that was standard practice for the district. The findings identified a positive effect of the CFC program on observed classroom text discussion quality. Supporting the theory underlying CFC, the positive effect of the program on student reading achievement was mediated through the quality of classroom text discussions. Students' language status moderated the direct effect of the program, with stronger effects for ELL students compared to their English-proficient peers.

1. Introduction

The ability to read with high levels of understanding is critical to academic success and to lifelong learning. While most children learn to decode texts and identify main ideas in the early elementary grades, many never advance beyond basic levels of comprehension (Buly & Valencia, 2002). This is especially the case for low-income, minority, and language-minority students who are over-represented within the lowest levels of reading skills in many countries (August, Carlo, Dressler, & Snow, 2005; Droop & Verhoeven, 1998; Lesaux, Rupp, & Siegel, 2007; Stuart, 1999). For example, in the United States only 19% of low-income fourth-grade students, and 8% of English language-learners (ELL) scored at or above proficient on a national assessment of students' reading skills meaning that they were able to recognize the main message of a text, connect ideas in a text, and apply information in a text to support an opinion (Lee, Grigg, & Donahue, 2007).

A considerable body of research shows that well-implemented classroom text discussions can be a key instructional practice for developing students' higher-level reading comprehension skills (Murphy, Soter, Hennessey, Alexander, & Nystrand, 2009; Nystrand, 2006). The rationale supporting the relation between text discussions and students' comprehension skills is derived in large part from sociocultural theory which foregrounds the role of social interaction in the development of higher-level mental processes (Vygotsky, 1986). From this perspective, interactive, high-level class discussions can develop comprehension skills by providing multiple opportunities for students to tackle challenging questions, test conclusions, attempt to reconcile multiple perspectives, and bring to fruition ideas that might otherwise have remained embryonic (Goldenberg, 1993; Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009). The social processes experienced by students in these discussions, for example, the thinking processes exhibited by a teacher and students as they seek to build meaning, ultimately emerge as mental functions within students that enable them to construct meaning from texts independently. Teachers nurture this development by guiding the discussions in a way that 'scaffolds' students' emerging understandings (Tharp & Gallimore, 1988).

The benefits of implementing interactive text discussions for improving reading comprehension are well documented (Murphy et al., 2009). Moreover, the positive effects of such discussions appear to be consistent across diverse populations of learners, including low-income and minority students (Applebee, Langer, Nystrand, & Gamoran, 2003). Research is still in the early phases of identifying best practices for improving the reading outcomes of low-income language-minority students who comprise a significant subgroup of students in many countries (Droop & Verhoeven, 1998;

* Corresponding author. Tel.: +1 011 1 412 624 6944.
E-mail addresses: lclare@pitt.edu (L.C. Matsumura), hgarnier@sasdatalink.com (H.E. Garnier), jessaca.spybrook@wmich.edu (J. Spybrook).
[1] Tel.: +1 011 1 412 624 6944.
[2] Tel.: +1 011 1 269 387 3889.

0959-4752/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.learninstruc.2012.11.001

Kieffer, 2008; Stuart, 1999). Some studies suggest, however, that classroom discussions that foster opportunities for active meaning making of rigorous academic content also are effective for increasing the reading comprehension skills of ELLs (Goldenberg, 1993; Saunders & Goldenberg, 2007).

Research indicates as well that discussions that encourage active participation and meaning making on the part of students rarely occur in classrooms (Applebee et al., 2003). Such discussions are a significant departure from recitation style classroom discourse characterized by teachers posing a series of close-ended questions that require students to display their mastery of a text through convergent factual answers followed by teacher evaluation or feedback (Cazden, 2001). Indeed, recitation style discourse is so prevalent that it has been described as a *de facto* script for classroom interactions (Tharp & Gallimore, 1988). Implementing high-level interactive text discussions is difficult for many teachers, in part because such discussions require teachers to teach in ways that they likely did not experience as learners in the school system or even in their teacher education program. Leading high-level text discussions also requires substantial skill and planning on the part of teachers. Teachers must know when to challenge students, to draw out a student's idea, to keep quiet, to clarify an idea from the text, and perhaps most importantly, to weave individual participants' comments into a "larger tapestry of meaning" (Goldenberg, 1993, p. 318).

Based on the belief that traditional forms of professional development available to teachers can be inadequate for helping teachers implement new and challenging forms of classroom discourse (Hawley & Valli, 1999), local education authorities (e.g., school districts) have turned to other forms of professional development that include hiring more skilled teachers to serve as literacy coaches. Literacy coaching offers a strategy for introducing into schools the characteristics of effective professional development gleaned from several lines of research (see studies reviewed in Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009). While different models of coaching exist (e.g., Biancarosa, Bryk, & Dexter, 2010; Costa & Garmston, 2002), the general consensus among models is that literacy coaches provide long-term engagement of teachers in a learning community with opportunities to design, enact, and critique lessons. Coaches also provide the guidance of a more expert practitioner who can teach the new content knowledge teachers need and is skilled in assisting teachers in their initially hesitant performances of new and demanding forms of teaching (Walpole & McKenna, 2004). When well implemented, coaching creates an apprenticeship environment (Collins, Brown, & Hollum, 1991) for learning the knowledge, skills and self-reflective strategies of the effective literacy teacher.

Literacy coaching for teachers has been widely endorsed. Notably, literacy coaching has been included in U.S. education reform policies at the district, state, and federal levels. Even though many studies of coaching show positive results on teaching and/or learning (e.g., Joyce & Showers, 1988; Teemant, Wink, & Tyra, 2011; Veenman & Denessen, 2001), until recently few studies existed that employed experimental designs to test the effect of school-based literacy coaching programs on teaching and learning relative to other forms of professional development. Moreover, while change in teaching quality is the presumed causal agent of literacy coaching's effect on student learning, no study to our knowledge has tested this assumption in an efficacy trial. In the current study, we address this gap in the research and present results from a group-randomized trial investigating the effect of Content-Focused Coaching (CFC) on classroom text discussion quality and students' reading achievement in schools serving a high proportion of minority and English language learning students from low-income families. Specifically, we examine the extent to which classroom text discussion quality mediates the effect of CFC on students' reading achievement.

1.1. Experimental and quasi-experimental studies of literacy-coaching programs

In this section, we review studies that have applied experimental and quasi-experimental designs to investigate the effect of literacy coaching on teaching behaviors and students' literacy outcomes. We identified only three studies that measured coaching's effect on both teaching and learning outcomes in the same design (Garet et al., 2008; Powell, Diamond, Burchinal, & Koehler, 2010; Sailors & Price, 2010). These three studies showed mixed results. Sailors and Price (2010) identified positive effects of coaching on both teachers' use of cognitive reading strategies and students' reading skills relative to teachers' participation in a professional development institute only. Powell et al. (2010) also showed positive effects of a coaching intervention, delivered onsite or remotely, on the structural features of the classroom literacy environment (e.g., how teachers organized writing centers), and positive effects of coaching on preschool children's knowledge of letters and print, blending skills, and writing. The coaching intervention did not impact the quality of teacher–student interactions or children's oral language development. Garet et al. (2008) found no effect of literacy coaching on teaching or on students' reading achievement relative to other forms of professional development. However, the training and support provided to coaches varied across these studies, which could possibly explain their mixed results. In both Sailors and Price (2010) and Powell et al. (2010), the coaches were either part of or closely associated with the research team. In the larger study conducted by Garet et al. (2008), coaches received relatively minimal training and support.

In the set of studies that focus solely on teacher outcomes, Neuman and her colleagues identified a positive effect of coaching on the quality of early language and literacy practices in center- and home-based care settings (Neuman & Cunningham, 2009). Stanulis, Little, and Wibbens (2012) examined the effect of a coaching intervention for new teachers in the elementary grades relative to standard practice for teacher mentoring in the district. They also identified a positive effect of coaching on teacher–student interactions, specifically, the quality of classroom text discussions. Since these studies were limited to teaching outcomes, however, it is not possible to evaluate if the observed changes in practice identified by the researchers were sufficient to increase students' literacy skills.

In the set of studies assessing the effect of coaching only on students' literacy skill development, Biancarosa et al. (2010) identified increasing positive effects of a comprehensive coaching program (the Literacy Collaborative) on primary grade students' reading skills. Lovett, Lacerenza, De Palma, Benson, Steinbach, and Frijeters (2008) also identified a positive effect of coaching on high school students' reading achievement, though only in the year following a coaching intervention suggesting a lagged effect. Van Keer and Verhaeghe (2005), in contrast, compared an intensive coaching model to a traditional in-service model implemented with second- and fifth-grade teachers and found that both interventions showed positive effects. In other words, coaching was no more effective for inducing change in student learning than a more traditional teacher professional development institute. None of these studies, however, examined how teachers' instruction was also influenced by the coaching they received. A limitation of these studies is that they do not provide insight into *how* coaching achieved its desired effects on student learning. In other words, they did not provide evidence of the key processes underlying the effects of coaching programs on student outcomes.

1.2. Content-focused coaching

The CFC program in elementary literacy was developed at the University of Pittsburgh's Institute for Learning (IFL) (Staub, West, &

Bickel, 2003). CFC is a multi-level intervention based on the theory that substantive improvement in student learning requires learning opportunities for actors across the levels of the school system (Resnick & Spillane, 2006). The program thus seeks to effect change in student achievement by fostering learning opportunities at multiple levels: district, school, and classroom.

At the district level, IFL fellows provide intensive professional development to literacy coaches to ensure they have a high level of pedagogical expertise and ability to work effectively with teachers in their classrooms and in professional learning groups. District leaders and principals in the CFC schools also participate in the coach professional development to help create a shared vision of the coaches' intended work with teachers. Research shows, for example, that the responsibilities of a literacy coach often vary widely across schools (Blarney, Meyer, & Walpole, 2009; Duessen, Coskie, Robinson, & Autio, 2007). The CFC program thus seeks to create consistency in how effective teaching and the coaching role is defined to help create the organizational conditions that support high-quality literacy coaching.

CFC-trained literacy coaches, in turn, return to their school and work with teachers in professional learning groups and individually in their classrooms. Working with teachers across classrooms is intended to help create the school-level conditions that support effective teaching. These school-level conditions include a vision of effective text discussions that is shared across teachers and is supported by the principal, as well as a collaborative culture where teachers support each others' learning and implementation of new instructional practices (McLaughlin & Talbert, 2006). CFC coaches work individually in teachers' classrooms to provide 'at-elbow' guidance for teachers to implement high-quality, interactive classroom discussions that increase students' reading comprehension skills.

In the following sections, we describe in more detail the substance and form of the professional development provided to CFC coaches and leaders to ensure coaching quality, and to teachers to improve the quality of their classroom text discussions.

1.2.1. CFC professional development activities for coaches

The subject-matter explicitness of CFC distinguishes this approach from many other coaching programs that focus on general coaching and teaching strategies (e.g., Costa & Garmston, 2002; Joyce & Showers, 1988). Specifically, the substance of the training provided to the CFC coaches and teachers centers on Questioning the Author (QtA), a discussion-based approach to reading comprehension (Beck & McKeown, 2006). The coaches and teachers in our study did not have prior experience enacting QtA as this pedagogical approach is not generally taught in teacher education programs or in traditional district-sponsored professional development. The CFC coach professional development focuses on ensuring that coaches have the knowledge and skills necessary to plan and model high-quality QtA lessons for teachers. QtA is the substance of the CFC intervention based on research linking this method to improved reading comprehension skills for students in the upper elementary grades (McKeown, Beck, & Blake, 2009), and based on the IFL's special expertise in this method.

QtA draws on cognitive science research that characterizes text comprehension as an active and inferential process of building a mental representation of a situation described by a text (e.g., Kintsch & Van Dijk, 1978). To develop students' skills at engaging in this process, QtA encourages teachers and students to work together to construct meaning of a text during the reading process. Teachers strategically pose questions to students at key places in a text that promote understanding, interpretation and elaborated response, and encourage students to share and challenge each other's ideas to grapple with these questions (Beck, McKeown, Sandora, Kucan, & Worthy, 1996). Because less-than-coherent texts can pose enormous comprehension problems for readers, QtA also encourages students to view authors as fallible writers who do not always express their ideas clearly (Beck & McKeown, 2001). The idea is to empower students to understand that the difficulties they encounter in comprehending a text could be a result of poor writing as opposed to deficits in their reading skills and thus embolden students to "question text ideas and dig into their meaning" (Beck et al., 1996, p. 387).

The form of the professional development provided to coaches to build their pedagogical skills at enacting QtA draws on a cognitive apprenticeship model of instruction (Collins, Brown, & Holum, 1991). The IFL fellows study the theory underlying QtA with coaches, share lesson plans that are rooted in the district's curricula, and model QtA lessons for coaches in teachers' classrooms. The coaches, in turn, co-plan QtA lessons with the IFL fellows and their coach colleagues and are observed teaching QtA lessons in teachers' classrooms before moving on to independently plan and model QtA lessons with teachers at their school. Following a similar pattern, coaches develop their coaching skills by first watching videotapes of IFL fellows conferring with individual teachers and leading teacher professional learning groups. The coaches then are observed conferring with teachers individually and leading professional learning groups (i.e., grade-level teams) and receive feedback from the IFL and their coach colleagues.

To monitor the quality of the coaches' pedagogical skills, coaches are observed twice teaching a QtA lesson in a teacher's classroom (once by the IFL fellows and once by a coach colleague). To monitor the quality of coaches' developing coaching skills, coaches also are observed twice (once by the IFL fellows and once by coach colleagues) leading a teacher learning group and conferring individually with teachers. CFC coaches also keep logs to track and analyze their use of time, and create portfolios of their work demonstrating their efforts in four key areas: 1) teaching students as a model for teachers; 2) leading teacher learning groups; 3) conferring individually with teachers; and 4) meeting regularly with their principal.

1.2.2. CFC professional development activities for teachers

The substance of the CFC coaches' work that they undertake with teachers focuses on increasing teachers' pedagogical knowledge and ability to plan, teach, and reflect on their lessons, with a specific emphasis on QtA. As described earlier, QtA techniques are not widely known by practitioners and teachers generally have significant difficulty planning for, and enacting, high-quality classroom text discussions (Applebee et al., 2003). Planning for high quality QtA lessons is a complex process that involves multiple skills on the part of teachers including the ability to evaluate texts for their potential to support high-quality discussions. CFC coaches teach teachers to consider both the level of coherence and 'grist' in the texts provided by their district based on research indicating that both features of texts (coherence and engaging content) play a role in students' comprehension (Beck & McKeown, 2001).

Coaches also work with teachers to develop queries that support comprehension. This includes supporting teachers to develop open-ended questions that support students to consider and extend meaning from a text and to make connections between the content of texts and larger concepts (Beck & McKeown, 2001, 2006). Coaches also teach teachers how to sequence questions in a way that supports students to make connections among ideas in the text as they read. In learning how to plan for QtA lessons, coaches work with teachers to select appropriate stopping places in a text for posing questions (Beck & McKeown, 2006). These can be places in a text that might be difficult for students to comprehend, for example, places where the author does not express themselves clearly or uses vocabulary that is likely to be unfamiliar to students. An appropriate stopping point also can be a place in a text where

students would need to activate, or be provided with, background knowledge necessary for comprehending that portion of a text (Beck & McKeown, 2006).

In addition to helping teachers develop high-level queries that focus attention on specific text content, coaches also work with teachers to facilitate more interactive text discussions. As described earlier, interactive and collaborative discussions are a significant departure from traditional forms of classroom discourse and are often very difficult for teachers to implement (Tharp & Gallimore, 1988). To support teachers' skills at facilitating such discussions, coaches teach strategies for connecting students' expressed ideas in discussions to show how their ideas relate to one another, encouraging students to talk to each other, and pressing students to explain and support their claims (Beck & McKeown, 2006).

To achieve these learning goals, coaches meet with teachers in weekly grade-level team meetings to study the theory underlying QtA, and meet with teachers on a monthly basis in their classrooms to engage in a cycle of three activities: lesson planning, enacting, and reflecting on instruction. Coaches begin their one-on-one work with teachers by modeling a QtA lesson in their classroom. In addition to showing teachers what QtA looks like in practice, beginning work with teachers in this way helps to establish the coach's credibility as an instructional specialist and to foster a learning culture in the school where everyone's practice – beginning with the coach – is put forth for critique. After this initial modeling phase, coaches observe teachers enacting QtA lessons, and/or co-teach lessons with teachers as needed. Coaches then reflect on the lesson with teachers with an emphasis on student learning evidenced in the discussions. Specifically, to help teachers apply a 'learner-centered' lens for reflecting on their instruction, coaches use a CFC-developed protocol (Staub, 2004) that guides teachers to think about the specific intended learning for a lesson, what students do and do not seem to comprehend, and how students with differing levels of skills might need to be assisted to learn the lesson content.

1.3. Research questions and hypotheses

In the present study, we address the following research questions: (a) Does classroom text discussion quality mediate the effect of CFC on students' reading achievement? (b) Is the effect of CFC on reading achievement moderated by students' language status? To answer the first question, we examine the direct and indirect effects of CFC to test the hypothesis that CFC will show a positive effect on the quality of the classroom discussion quality (Hypothesis 1a) and that the impact of the program on students' reading achievement will be mediated by the quality of classroom text discussions (Hypothesis 1b). As described earlier, we are not aware of any other study that has specifically tested the theory of action that literacy coaching improves student learning through improved teaching quality.

To answer the second question, we explore the moderating effects of students' language status on the direct and indirect effects of CFC. Research evidence is insufficient to support a formal hypothesis for this question. We include this question in our study because of the growing numbers of ELL students in the United States (Kindler, 2002), and the difficulty teachers often have meeting their learning needs (Kelley, Lesaux, Kieffer, & Faller, 2010). It is important to identify professional development interventions, such as CFC, that can be effective for improving the reading achievement of ELLs who often are educated with their non-ELL peers by the same teacher.

The final purpose of our study is to contribute to the literature on the use of multilevel mediation analyses in randomized field trials in education settings. Mediation analyses are commonly applied in other fields to identify mechanisms through which interventions achieve their effects such as health related interventions (Stigler, Kugler, Komro, Leshabari, & Klepp, 2006), and in psychological research (MacKinnon, Fairchild, & Fritz, 2007) and epidemiological studies (VanderWeele, 2009). Fewer applications are found in experimental studies of educational interventions when three levels are present – students, classrooms, and schools – in part because this statistical technique is not widely known (Pituch, Murphy, & Tate, 2010; Raudenbush, 2011). Our group randomized trial with random assignment at the school level, the mediator (text discussion quality) at the classroom level, and the outcome (reading achievement) at the student level provides an opportunity to explore and illustrate the types of information yielded by this analytic approach when applied in an education setting. We believe that this approach to analysis could be useful for studying other school-based interventions (e.g., particular comprehensive reform models) that aim to increase student learning through improved teaching quality (Cohen, Raudenbush, & Ball, 2003).

1.4. Design of the trial

To examine the effectiveness of the CFC program for improving teacher and student outcomes, we conducted a group-randomized trial in which schools within one district were assigned to the CFC program or comparison condition. Schools assigned to the treatment condition received a CFC-trained coach and schools in the comparison condition continued with the literacy coaching that was standard practice in the district. We define standard practice as professional development activities a school would normally engage in without interference or input from the study's researchers. District-mandated curricula were in place prior to our study and were similar across conditions in order to ensure that treatment and comparison school differences in outcomes were not confounded with curricula quality. CFC coaches averaged 12 years of elementary teaching experience (ranging from 0 to 22 years) and 2 years of prior coaching experience (ranging from 0 to 9 years). Literacy coaches in the comparison schools averaged 15 years of prior teaching experience (ranging from 4 to 40 years) and three years of prior coaching experience (ranging from 1 to 6 years) (see Matsumura, Garnier, & Spybrook, 2012, for a further description of the literacy coaches in both conditions).

The natural clustering in schools, with students nested within teachers nested within schools, and implementation at the school level make a group randomized trial the most appropriate design for studying program effectiveness (Bloom, 2005; Boruch & Foley, 2000). As long as the nested structure of the data is accounted for in the design and analysis of the trial, a group randomized trial yields unbiased estimates of the treatment effect (Raudenbush, 1997). We chose to randomize at the school level rather than the classroom level to minimize potential dilution of treatment effects that could occur within schools if teachers in the treatment condition discussed and shared CFC content and materials with teachers in the control condition. The specific design of the CFC program also supports randomizing at the school level because an important goal of the program is to create school-based teacher learning groups and coach-principal partnerships. Randomizing at the teacher level would undermine the implementation of the CFC intervention, and pose a significant burden on principals to support the coaching intervention for some teachers but not for others.

1.4.1. Timeline of the trial

The CFC program was implemented in schools for three years, consistent with the literature that suggests it can take several years for multi-level interventions to show effects on desired outcomes (Fixsen, Blase, Naoom, & Wallace, 2005). CFC-trained coaches were introduced to the treatment schools relatively late in the first study year. According to coach interviews, the large focus of the coaches' work in the first year was establishing relationships with their

schools' principals and gaining teachers' trust (Matsumura, Garnier, & Resnick, 2010). In the second year of the trial, the coaching intervention was more fully established in schools and coaches began working with teachers in August. Interim effects were investigated at the end of year 2 in order to monitor the implementation of the program and also to assess the progress of the program.

2. Methods

2.1. Sample

2.1.1. Schools

A sample of 32 schools serving primarily minority and English language learning (ELL) students from low-income families were originally recruited for this trial. All schools were from a medium-sized district in the Southwestern United States. Schools were randomly assigned to either the CFC or comparison condition in the summer before the trial was implemented. For transparency, random assignment was conducted in a face-to-face meeting with the researchers, district administrators and principals. At this meeting, a principal volunteer drew names of schools and alternately assigned these to treatment and comparison conditions. One intervention and two comparison schools withdrew from the trial after assignment and before implementation of the program for reasons of high teacher workloads and lack of interest in the trial. The final sample included 29 schools (15 intervention and 14 comparison).

2.1.2. Teachers

The study design called for all of the fourth- and fifth-grade teachers in each of the 29 schools to participate in the research activities. These grades were targeted for the intervention based on research showing that low-income students' reading skills often begin to decline in the fourth grade relative to their wealthier peers (Chall, Jacobs, & Baldwin, 1990). Of the 203 eligible teachers, 192 agreed to participate (95%) in the study at the beginning of year 1 (94% and 95% of the eligible teachers in the CFC and comparison schools respectively). The reason teachers in both conditions identified for not participating in the study was that they did not have time to complete the research activities. On entry to the study, teachers in the CFC and comparison schools did not differ on background characteristics (highest degree earned, certificate status, teaching a designated ELL class, years teaching, years teaching reading), past professional development participation including exposure to literacy coaching, perception of the quality of their principals' leadership, perception of their school's norms for professional collaboration, perception of trust and respect among their teacher colleagues. By the end of the year, 92% ($N = 177$) of our participating sample remained in the study with no difference in attrition rates across conditions. The reasons teachers left the study were maternity leave, leave of absence for personal reasons, and departmentalization (i.e., the teacher no longer taught reading). Two teachers, both from comparison schools, left the study because they no longer wanted to participate.

At the beginning of the second year, 98 teachers remained in the study (61% treatment and 39% comparison). Such a large teacher attrition rate (45%) from our sample schools after the first year is unusual, but not wholly unexpected. Henusheck, Kain, and Rivkin (2004) showed that teacher attrition is much higher in schools with large numbers of low-income and minority students than in schools serving more privileged students. We compared treatment and control groups on attrition rates and found no significant association between attrition rates and group assignment. In addition, we compared the background characteristics of teachers who returned after year 1 and those who left after year 1. No differences were detected. In fact, nearly all of the teachers who left the study did so because they left their school.

Since a goal of the CFC program is to create the school-level conditions that support instructional improvement by engaging all teachers within a grade and encouraging teachers to work collaboratively in grade-level teams with their coach, we recruited the teachers who were hired to replace those that left their school at the end of year 1. Seventy-three replacement teachers were added to the sample (92% of the eligible teachers). The reason teachers gave for not participating was that they were too busy to complete the research activities. Upon entry into the study in year 2, the replacement teachers in the CFC and comparison schools did not differ on background characteristics, and past professional development participation including exposure to literacy coaching. Because the teachers were new to their schools, we did not include their perceptions of their school's leadership and teacher relationships in our baseline comparisons.

We also compared the background characteristics of teachers who remained after year 1 and those newly recruited in year 2. Results indicated the newly recruited teachers had significantly fewer years experience teaching in general and were more likely to hold temporary certification. These measures of pre-existing differences will be included as teacher-level covariates in subsequent analyses as well as the number of years teachers participated in CFC to control for pre-existing differences in length of exposure to the program.

The teachers who comprise the final sample of our current study include participating teachers who entered the study in year 1 and teachers who were newly recruited in year 2. Although we collected observation ratings in the fall and spring of year 2 for the full final sample of 171 teachers, our analyses include those teachers for whom we have *both* observation ratings of their classroom discussions and student reading achievement scores on the state assessment test ($N = 167$). No differences in background characteristics or baseline student reading achievement were detected between the 101 CFC teachers and 66 comparison teachers (Table 1). In general, the 167 teachers were all full-time teachers. Thirty-three percent of the teachers had attained a Master's degree, 14% held temporary, provisional, or emergency certification, and 2% were certified through the National Board for Professional Teaching Standards. The teachers reported a wide range of teaching experience, ranging from 0 to 35 years and averaging 8 years teaching generally and 8 years teaching reading (median of 5 years teaching).

2.1.3. Students

The district provided student background information for $N = 2983$ students in this study. No differences between CFC and comparison schools were detected on student characteristics or baseline reading achievement (see Table 2). Overall, the large majority of these students qualified for free or reduced-price lunch (91%). Students were mostly Hispanic (79%) or African American (16%). Fifty-six percent were female. Forty percent of the students were designated as English language learners (ELLs) by the school district, and nearly all of these students were native Spanish speakers.

Forty-one percent of the ELLs took the state accountability test in English. The remaining ELL students took the state accountability test in Spanish. We collected this information because it provides some insight into ELL students' language status beyond that which is provided by the ELL designation, specifically with regard to students' progress toward English proficiency. ELL students in our study's district take the state accountability test in English when teachers believe that they have mastered the language well enough to be successful. ELL students who pass the state test in English for two consecutive years are re-designated by the school district as non-ELLs. We acknowledge that neither the measure of ELL status nor language version of the state test provides precise information

Table 1
Baseline survey reports of teacher background information, quality of text discussion rating, and student reading achievement for 167 teachers in the study through year 2.

Background measure	CFC (n = 101 teachers)		Comparison (n = 66 teachers)	
	M	*SD*	*M*	*SD*
Bachelors degree (%)	100	0	100	0
Masters degree (%)	34	48	31	47
Certified (%)	98	11	98	15
Regular certificate (%)	81	39	81	40
Temporary certificate (%)	16	37	11	31
National Board certificate (%)	1	11	4	20
Teaching a designated ELL class (%)	60	49	60	50
Teaching in general (years)	8	8	9	10
Teaching reading (years)	7	8	8	10
S7 English language version of test (%)	68	41	72	38
S7 state standardized reading achievement (score)	2196	68	2191	67

Note. Baseline survey responses were provided by 126 of the 167 teachers remaining in the study. No background differences were detected between CFC and comparison teachers.

about students' language development in either Spanish or English, but we believe that together these two variables provide critical information about the language status of students in our sample.

2.2. Instruments and procedures

2.2.1. Quality of classroom text discussion ratings

Teachers were observed holding a classroom text discussion in the fall (September/October) and spring (April/May) of each study year (n = 2 observations per year). In year 1 of the study, observations were conducted prior to the introduction of the CFC coaches to schools in November. The observation component of the Instructional Quality Assessment (Matsumura, Garnier, Slater, & Boston, 2008) was used to measure the quality of classroom text discussions with seven ratings (see Table A.1). Seven observers were recruited from a local university's School of Education and through an advertisement posted on a public website (*craigslist*). The observers participated in a total of five days of training that consisted of studying the IQA rubrics, rating lessons from videotaped classroom discussions, and practicing field note writing. The data collection supervisor for the project also observed with the individual observers in the sample classrooms to monitor reliability for two to four lessons with each observer depending on scheduling. Observers who were not able to obtain at least 80% agreement with the data collection supervisor over two consecutive lessons did not continue in the study. Observers met to review the rubrics and reliability was re-checked prior to each data collection period. Inter-rater agreement also was checked again midway through each data collection period to help guard against 'drift' in the ratings. Overall agreement between the observers and the data collection supervisor across all observations for each rater pair averaged 91% (fall) 84% (spring) of year 1 of the trial, and 86% (fall) and 87% (spring) of year 2.

Observers requested in advance to visit teachers' classrooms during a 45-min portion of their language arts block when the class held a text discussion. Observers sat at the back of the class and took extensive field notes during the observation in order to record as much of the teacher and student interaction as possible. The rating scales were completed outside of the classroom immediately following each observation. At each observation, the large majority of teachers (over 90%) reported that the observed discussions were representative of the text discussions that normally occurred in their classroom. Observations were conducted in the language of instruction that was typical for the teacher (two of the observers were bilingual). Consistently across both study years most of the observed text discussions (approximately 75%) were conducted in English, 20% were conducted in a mixture of English and Spanish, and 5% were conducted in Spanish only.

The IQA is constructed around principles of Accountable Talk (Michaels, O'Connor, & Resnick, 2008), a framework that draws on both sociocultural theory and research in the learning sciences (e.g., Bransford, Brown, & Cocking, 2000). In this study, seven ratings from the IQA were averaged to create one baseline and one spring 2008 index measuring the quality of teacher and student participation in the discussion about a text and the academic rigor of the discussion (α = .78 based on the sample in this study). Each rating was assessed on a four-point scale (coded 1 = 'Poor' to 4 = 'Excellent'): 1) Widespread student participation in the discussion; 2) Teacher connects students' ideas to build coherence in the discussion; 3) Students make contributions that link to and build on each other; 4) Teacher presses students to explain their answers or support their claims with evidence from a text; 5) Students explain and support their claims; 6) Academic rigor of the text used in the discussions; and 7) Academic rigor of the discussion with the teacher supporting the students to analyze and interpret a text. No baseline difference was detected between teachers in CFC and comparison schools in the quality of their observed classroom text discussions.

Table 2
Baseline school characteristics of student demographics and reading achievement.

School measure	CFC (n = 15 schools)		Comparison (n = 14 schools)	
	M	*SD*	*M*	*SD*
Grade 4 (%)	61	15	51	17
Gender				
Female (%)	52	4	50	5
Ethnicity				
African American (%)	12	9	21	19
Hispanic (%)	83	12	75	21
White (%)	4	6	4	5
Other (%)	1	4	1	1
Qualified for free lunch program (%)	91	8	91	7
Immigrant status (%)	4	3	3	2
Limited English proficient status (%)	44	15	33	21
S7 English language of state standardized test (%)	68	16	78	19
S7 state standardized reading achievement (score)	2198	29	2188	47

Note. No baseline differences were detected on any school-level measures.

2.2.2. Student reading achievement

Students' reading achievement was assessed using the state standardized accountability test in years 1 and 2. Items on the reading portion of the test are designed to assess students' grasp of the basic events in a text as well as their higher-level comprehension skills. The items are grouped under four main objectives for students: 1) demonstrate a basic understanding of culturally diverse written texts (including reading/vocabulary development, ability to determine a text's main idea, or summarize a text); 2) apply knowledge of literary elements to understand culturally diverse written texts (including the ability to analyze characters and recognize and analyze story plot, setting, and problem resolution); 3) use a variety of strategies to analyze culturally diverse written texts (e.g., use the text's structure or progression of ideas to locate and recall information, judge the internal consistency or logic of stories); and 4) apply critical-thinking skills to analyze culturally diverse written texts (e.g., draw inferences such as conclusions or generalizations and support them with text

evidence). The test contains both fiction and non-fiction passages. The overall reliability of this test is high ($\alpha = .88$ at fourth grade and .87 at fifth grade) (Technical Digest, n.d.).

The scores from the state accountability test are not vertically scaled, implying that a student's scale score in one grade cannot be compared to the student's scale score in the next grade.[3] Table 3 presents the means and standard deviations for year 2 for the full sample of $N = 2983$ students, by ELL designation and grade level.

2.2.3. Fidelity of CFC program implementation

Teachers (93 CFC and 45 comparison) completed surveys at the end of the second year of the study describing their interactions with their school's literacy coach. Teachers described how often they met with their coach individually and in grade-level teams meetings, and how often coaches modeled lessons in their classroom, observed them teaching, or co-taught a lesson with them (coded 1 = 'Never', 2 = 'One to three times a year', 3 = 'Four to six times a year', 4 = 'Monthly' and 5 = 'Weekly'). The content emphasized in the coaching activities in which teachers participated was assessed with 16 items (coded 1 = 'None', 2 = 'Minor emphasis', 3 = 'Moderate emphasis', and 4 = 'Major emphasis') and averaged into four dimensions: 1) planning and reflecting on instruction (e.g., my coach modeled for me how he/she plans and reflects on a lesson; my coach improved my ability to anticipate student difficulties or misconceptions), 2) enacting lessons in their classrooms (e.g., my coach helped me facilitate better classroom discussions; my coach helped me determine what to teach next), 3) building knowledge of the theory underlying effective pedagogy (e.g., my coach studied educational articles with me, and 4) differentiating instruction (e.g., my coach helped me adapt lessons to specific student needs) ($\alpha = .80$ to .88 based on the sample in this study).

2.3. Data analyses

Our main goal is to investigate if the effect of the CFC program on student achievement will be mediated by changes in quality of text discussion. Since we are interested in student level outcomes, the mediator is at the teacher level, and random assignment occurred at the school level, we use a 3 → 2 → 1 mediation design (Pituch et al., 2010). The basic mediation model is depicted in Fig. 1 where a represents the effect of CFC on quality of classroom text discussion, b represents the effect of quality of classroom text discussion on student reading achievement, and c' represents the direct effect of the CFC program on reading achievement. Given this model, the total effect is:

$$\begin{array}{ccccc} \text{Total effect} & = & \text{Indirect effect} & + & \text{Direct effect} \\ c & = & ab & + & c' \end{array} \quad (1)$$

However, our model is slightly more complicated because we also explore whether ELL status moderates the direct and indirect effects of CFC on student outcomes. Hence, the total effect, path c, the direct effect, path c', and the indirect effect, path ab, would be moderated effects. Because ELL is binary, we calculate the total moderated effect as a weighted average of the treatment effect for non-ELL and a treatment effect for ELL where the weights are the proportion of ELL and non-ELL students in the sample.

We used HLM 6.0 (Raudenbush & Bryk, 2002) to run a series of models to estimate each effect. The models we used to estimate each effect are included in the results section. To test the indirect effect, we followed the procedure described by Pituch et al. (2010).

[3] According to the state education agency scores on the vertical scale were made available on student reports for the first time in spring 2009, subsequent to the CFC study data collection.

Table 3
Means and standard deviations of the year 2 student reading achievement scores on the state accountability test.[a]

Sample of students	School group	Year 2 state accountability test	
		M	*SD*
All students	CFC schools	2178.79	168.39
	Comparison schools	2169.78	178.98
ELL students	CFC schools	2149.11	164.02
	Comparison schools	2104.04	162.57
Grade 4			
All students	CFC schools	2169.27	174.70
	Comparison schools	2155.69	183.70
ELL students	CFC schools	2144.27	173.81
	Comparison schools	2093.41	166.35
Grade 5			
All students	CFC schools	2194.08	156.63
	Comparison schools	2185.52	172.34
ELL students	CFC schools	2158.35	143.28
	Comparison schools	2118.34	156.68

Note. $N = 2983$ students.
[a] The state accountability test is administered each spring. The test was not a vertically-aligned assessment instrument across grades when the study data were collected.

We constructed confidence intervals using the Sobel (1982) standard error, $\sqrt{a^2 o_b^2 + b^2 o_a^2}$ where o_a and o_b are the standard errors of the a and b paths. We used the PRODCLIN software program (MacKinnon, Fritz, Williams, & Lockwood, 2007) to test the significance of the indirect effect.

3. Results

3.1. Direct and indirect effects

3.1.1. Path c: total effect of CFC on student reading achievement

We estimated the total effect (c in equation (1)) of the CFC program on student reading achievement using a three-level model. Since we are also interested in whether ELL status moderated the overall effect of CFC, we included ELL status at the student level and its interaction with the CFC program. In addition to ELL status, we also included the English language version of the test at the student level. To account for grade level differences in the sample, we included a grade level designation for Grade 4 or Grade 5 in our analyses. Both grades are included together in our model since our study is estimating the school-level effects of the CFC program for multiple grades. The year 1 test scores (baseline) are correlated with the year 2 test scores ($r = .65$, $p < .01$), hence we use the year 1 scores aggregated to the school level as a covariate in the analyses to improve the precision of the estimate of the treatment effect (Raudenbush, Martinez, & Spybrook, 2007). All level-1 covariates are grand mean centered and fixed at level-2 except the ELL gap, which we allow to vary at level-2. The student level, or level-1, model is:

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{ELL}) + \pi_{2jk}(\text{Grade}) + \pi_{3jk}(\text{English version}) + e_{ijk} \quad (2)$$

where Y_{ijk} is the reading score for student i in class j in school k, π_{0jk} is the adjusted mean reading score in class j in school k, π_{1jk} is the ELL gap in class j in school k, π_{2jk} through π_{3jk} are the other covariate effects, and e_{ijk} is the person-level error.

The teacher level, or level-2 model is:

$$\begin{array}{l} \pi_{0jk} = \beta_{00k} + r_{0jk} \\ \pi_{1jk} = \beta_{10k} + r_{1jk} \\ \pi_{2jk} = \beta_{20k} \\ \pi_{3jk} = \beta_{30k} \end{array} \quad (3)$$

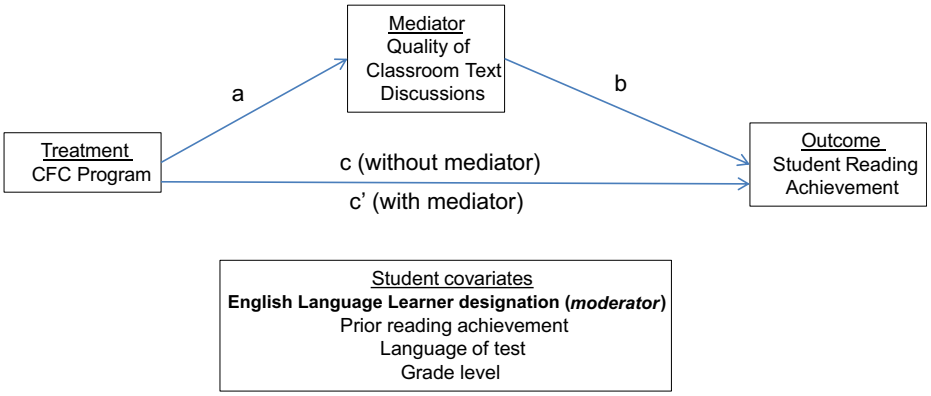

Fig. 1. Hypothetical mediation model of CFC effect on Student reading achievement through quality of classroom text discussions.

where β_{00k} is the adjusted mean for school k, β_{10k} is the ELL gap for school k, and β_{20k} and β_{30k} are the effects of the covariates for school k. The mean and ELL gap are allowed to vary randomly within schools.

The school level, or level-3, model is:

$$\begin{aligned}\beta_{00k} &= \gamma_{000} + \gamma_{001}(\text{CFC}) + \gamma_{002}(\text{Prior achievement}) + \mu_{00k}\\ \beta_{10k} &= \gamma_{100} + \gamma_{101}(\text{CFC}) + \mu_{10k}\\ \beta_{20k} &= \gamma_{200}\\ \beta_{30k} &= \gamma_{300}\end{aligned} \tag{4}$$

where γ_{000} is the overall adjusted mean achievement in non-CFC schools, CFC is the treatment indicator (1 = CFC, 0 = non-CFC), γ_{001} is the treatment effect, γ_{002} controls for the prior achievement, γ_{100} is the ELL gap, γ_{101} is the ELL by CFC interaction, or the moderator effect. We allow the mean and ELL gap to vary randomly across schools. Because ELL moderates the overall effect, we calculate the overall moderated treatment effect, or path c, as a weighted average of the treatment effect for non-ELL students and the treatment effect for ELL students, or $p_1(\widehat{\gamma}_{001}) + p_2(\widehat{\gamma}_{001} + \widehat{\gamma}_{101})$, where p_1 is the proportion of non-ELL students and p_2 is the proportion of ELL students.

Table 4 presents the results. The significant interaction ($\gamma_{101} = 48.304$, $p = .032$) suggests that the treatment has a differential effect for non-ELL and ELL students. CFC had a positive effect for non-ELL students ($\gamma_{001} = 19.177$, effect size (ES)[4] = .18, $p = .048$). The treatment effect for ELL students was $19.177 + 48.304 = 67.481$, $ES = .46$. That is, on average, the adjusted mean for ELL students in CFC schools was 67.481 points higher than ELL students in non-CFC schools. Hence the overall moderated treatment effect is .6 (19.177) + .4 (67.481) = 38.498, where .6 and .4 represent the proportion of non-ELL and ELL students, respectively, and $ES = .29$.

3.1.2. Path a: effect of CFC on quality of text discussion

We examined the effect of the CFC program on end of year 2 quality of text discussion ratings, the mediator (path a in Fig. 1), using a two-level model with teachers nested within schools. Initially, we included teacher reports of number of years teaching, temporary certification status, and baseline quality of text discussions ratings. In addition, we included the number of years in the study to control for pre-existing differences found between the year 1 teachers remaining in the study and the newly-recruited year 2 teachers. Preliminary analyses indicated that none of the teacher-level covariates were significantly associated with the quality of text discussions outcome and were omitted in favor of a more parsimonious model. The teacher level, or level-1, model is:

$$M_{jk} = \beta_{0k} + r_{jk} \tag{5}$$

where M_{jk} is the observed rating for quality of text discussion for teacher j in school k, β_{0k} is the mean quality of text discussion in school k, and r_{jk} is the teacher-specific error.

The school level, or level-2, model is:

$$\beta_{0k} = \gamma_{00} + \gamma_{01}(\text{CFC}) + r_{u0k} \tag{6}$$

where γ_{00} is the mean quality of text discussion in non-CFC schools which varies randomly across schools and γ_{01} is the effect of the treatment on the mediator. We are interested in γ_{01} which corresponds to path a in Fig. 1.

From Table 5, we can see that CFC had a positive impact on the quality of text discussion ($\gamma_{01} = .314$, $ES = .63$, $p = .002$). These results are similar to Neuman and Cunningham (2009) who found a substantial effect size of .77 for coaching with coursework on early language and literacy practices relative to teachers who participated in coursework only or who served as a comparison sample.

3.1.3. Paths c′ and b: effect of CFC on student achievement and effect of quality of text discussion on student achievement

To examine the direct effect of CFC on student achievement, path c', and the effect of the mediator on student achievement, path b, we use a three-level model beginning with the same model we used to estimate the total effect of CFC on student reading achievement. Initially, we included two measures of the mediator, an individual teacher-level measure of quality of instruction and an aggregate of this measure at the school level, both grand-mean centered. Following Pituch et al. (2010), we included the aggregated school-level mediator to examine the contextual effect of quality of text discussions, that is, the impact of school quality of text discussions on students reading achievement after controlling for classroom quality of text discussions. However, preliminary analysis indicated that the aggregated mediator was not significant, and it was omitted in favor of a more parsimonious model. In addition, preliminary analyses suggested that the indirect effect was not moderated by ELL status. Hence, the models presented below do not include ELL status as a moderator of the indirect effect.

The student level model is identical to equation (2). We differentiate the models by adding a' symbol to each coefficient. The teacher level model is:

$$\begin{aligned}\pi'_{0jk} &= \beta'_{00k} + \beta'_{01k}\left(M_{jk}\right) + r'_{0jk}\\ \pi'_{1jk} &= \beta'_{10k} + r'_{1jk}\\ \pi'_{2jk} &= \beta'_{20k}\\ \pi'_{3jk} &= \beta'_{30k}\end{aligned} \tag{7}$$

[4] The effect size for HLM analyses with cluster-level assignment is the adjusted group mean difference divided by the unadjusted pooled within-group standard deviation as described by the What Works Clearinghouse (2008).

Table 4
Results of HLM analysis estimating a model without mediator: total effects of CFC on the year 2 state accountability test of Student reading achievement.

Effects	Coefficient	*SE*	*p*-value
Final fixed effects			
Mean TAKS score, γ_{000}	2169.259	4.537	.000
CFC effect on TAKS score, γ_{001}	19.177	9.275	.048
Prior achievement, γ_{002}	.401	.127	.004
ELL gap, γ_{100}	−58.046	13.245	.000
Differential effect of CFC on ELL gap, γ_{101}	48.304	21.341	.032
Grade level difference, γ_{200}	24.206	7.799	.002
English language of test, γ_{300}	22.385	11.644	.054
Final random effects	**Variance component**	**χ^2**	***p*-value**
Level −1 effect, e_{ijk}	26096.717		
Mean achievement (teacher level), r_{0jk}	1844.925	208.048	.000
ELL gap (teacher level), r_{1jk}	794.825	122.280	.004
Mean achievement (school level), u_{00k}	15.392	25.404	>.500
ELL gap (school level), u_{10k}	1061.760	43.792	.022

Note. $N = 2983$ students, $N = 167$ teachers, $N = 29$ schools.

where β'_{000} is the adjusted mean for school k,, β'_{10k} is the ELL gap for school k, β'_{20k} and β'_{30k} are the effects of the covariates for school k which are held fixed, and β'_{01k} is the effect of the mediator on student achievement for school k. We allow the mean and ELL gap to vary randomly within schools.

The school level model is:

$$\begin{aligned}\beta'_{00k} &= \gamma'_{000} + \gamma'_{001}(\text{CFC}) + \gamma'_{002}(\text{Prior achievement}) + \mu'_{00k}\\ \beta'_{01k} &= \gamma'_{010}\\ \beta'_{10k} &= \gamma'_{100} + \gamma'_{101k}(\text{CFC}) + \mu'_{10k}\\ \beta'_{20k} &= \gamma'_{200}\\ \beta'_{30k} &= \gamma'_{300}\end{aligned} \tag{8}$$

where γ'_{000} is the adjusted overall mean student achievement in non-CFC schools, γ'_{001} is the effect of CFC on mean student achievement, γ'_{002} controls for the prior achievement γ'_{010} is the effect of the mediator on student achievement, γ'_{100} is the overall ELL gap, γ'_{101} is the ELL by CFC interaction, or the moderator effect, γ'_{200} and γ'_{300} are the effects of the covariates which are held fixed across schools. We are interested in γ'_{010} which is path b and $p_1(\gamma'_{001}) + p_2(\gamma'_{001} + \gamma'_{101})$ which is path c'.

The results are presented in Table 6. The quality of text discussions was significantly and positively associated with reading achievement (path b) ($\gamma_{010} = 22.030$, $ES = .21$, $p = .013$). The direct effect of CFC on adjusted reading achievement for non-ELL students is ($\gamma_{001} = 12.904$, $ES = .12$, $p = .175$). The direct effect for ELL students is now 12.904 + 45.786 = 58.690, $ES = .44$. Hence, the moderated direct effect is .6 (12.904) + .4 (45.786) = 26.056, $ES = .25$. This effect size is consistent with the literature in this area. For example, the study of comprehensive coaching programs by Powell et al. (2010) revealed effect sizes ranging from .17 to .29, and the average value-added estimates of the impact of the Literacy Collaborative on student achievement (Biancarosa et al., 2010) ranged from .28 to .33.

Table 5
Results of HLM ANOVA analysis estimating a model of the mediator: quality and academic rigor of classroom discussions.

Effects			
Final fixed effects	**Coefficient**	***SE***	***p*-Value**
Mean of text discussion quality at Spring 2008, γ_{00}	2.201	.060	.000
CFC effect on mean text discussion quality at Spring 2008, γ_{01}	.314	.088	.002
Final random effects	**Variance component**	**χ^2**	***p*-value**
Level-1 effect, r_{jk}	.205		
Mean of text discussion quality at Spring 2008, u_{0k}	.024	45.713	.014

Note. $N = 167$ teachers, $N = 29$ schools.

3.1.4. Path ab: mediated effect of CFC on student reading achievement

Using equation (1), we now can examine the direct and indirect effects. The total indirect effect is .314*22.030, or 6.917. Using PRODCLIN, the confidence interval we obtained was (1.23, 14.60) suggesting a statistically significant mediation effect. In terms of the overall moderated effect of CFC on student achievement (38.498), approximately 18 percent (6.917/38.498) is mediated by quality of text discussion. The final direct and indirect effects are shown in Fig. 2.

3.2. Literacy coaching in the CFC and comparison schools

We examined teacher reports of their participation in coaching activities and the content emphasized in those activities to investigate the fidelity of implementation of the CFC program and to provide insight on how the coaching teachers received differed across conditions. Understanding what occurred in the actual implementation of the CFC program is necessary for us to more completely understand the observed effects of the CFC program and how they are consistent with the CFC program's theory of change and operational definitions. Identifying differences between the coaching teachers received in the CFC and comparison conditions also is important for understanding the program's effect, and is especially important in a study such as ours where the comparison condition of standard practice in a district is less well defined than what would be the case in a laboratory setting.

Coaches were intended by the CFC program developers to meet with participating teachers in weekly grade-level team meetings to study the theory underlying effective reading comprehension instruction, and to meet with teachers individually at least on a monthly basis to engage in a cycle of lesson planning, enacting, and reflecting on instruction. Teacher reports identified few teachers in CFC schools who participated in the coaching activities at the levels intended by the CFC program developers by the end of the second trial year (Table A.2). Just 38% of the teachers in the CFC schools reported that they met with their coach in grade-level teams on a weekly basis and only 43% of the teachers reported that they met individually with their coach once a month or more. Only 28% of the teachers were observed by their coach for at least

Table 6
Results of HLM analyses estimating a model with mediator: Direct and indirect effects of CFC on the year 2 state accountability test of Student reading achievement.

Effects	Coefficient	*SE*	*p*-value
Final fixed effects			
Mean TAKS score, γ'_{000}	2169.860	4.357	.000
CFC effect on TAKS score, γ'_{001}	12.904	9.255	.175
Prior achievement, γ'_{002}	.341	.130	.015
Effect of quality of instruction, γ'_{010}	22.030	8.720	.013
ELL gap, γ'_{100}	−57.799	13.026	.000
Differential effect of CFC on ELL gap, γ'_{101}	45.786	20.971	.038
Grade level difference, γ'_{200}	25.927	7.910	.001
English language of test, γ'_{300}	22.389	11.236	.046
Final random effects	**Variance component**	**χ^2**	***p*-value**
Level −1 effect, e'_{ijk}	26114.481		
Mean achievement (teacher level), r'_{0jk}	1749.810	207.604	.000
ELL gap (teacher level), r'_{1jk}	635.356	122.010	.004
Mean achievement (school level), u'_{00k}	14.308	25.313	>.500
ELL gap (school level), u'_{10k}	1062.939	45.348	.015

Note. $N = 2983$ students, $N = 167$ teachers, $N = 29$ schools.

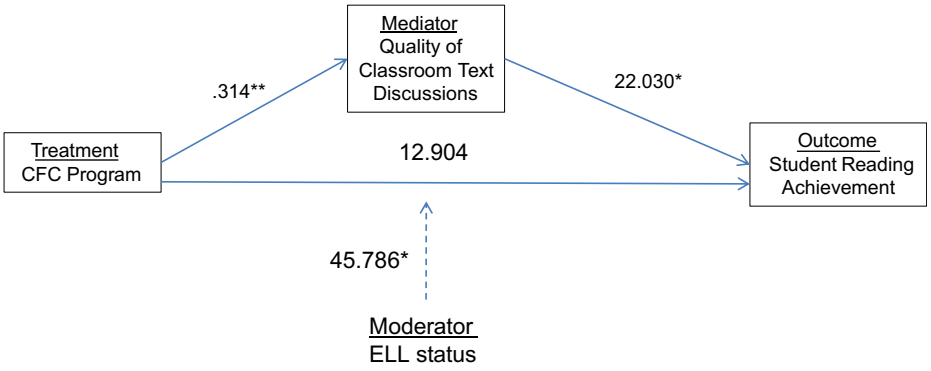

Fig. 2. Mediation model of CFC effect on Student reading achievement through quality of classroom text discussions with ELL moderating the direct effect $^{**}p < .01$; $^{*}p < .05$.

30 min on a monthly or weekly basis. Fewer teachers had a coach model a lesson in their classroom or co-taught a lesson with a coach on at least a monthly basis (21% and 13%, respectively). Only 16 teachers participated in *all* of the coaching activities at the level intended by the CFC program developers. Notably, however, many CFC teachers participated in coaching *near* the desired level of frequency. Seventy-four percent of the CFC teachers reported that they met with their coach in grade-level teams once a month or more and about two-thirds of the teachers (68%) reported meeting with coaches individually at least 4–6 times a year. Only 10% of the CFC teachers reported that their coach *never* observed them teaching or modeled a lesson in their classroom.

Teachers further reported that their coaching activities emphasized content aligned with the goals of the CFC program. At the end of year 2, the large majority of teachers in the CFC schools reported that the coaching they received moderately or strongly emphasized planning and reflecting on instruction (67%), providing help during lesson enactment (66%), building theory underlying pedagogy (48%), and differentiating instruction (56%) (Table A.3).

Although the coaching participation of the sample of CFC teachers did not reach the desired level of frequency, it was significantly higher than comparison teachers in four of the five activities: Meeting with coach individually ($t(134) = 2.82$, $p < 01$), coach observing teaching for at least 30 min ($t(135) = 4.70$, $p < .001$), coach teaching a model lesson ($t(134) = 5.87$, $p < .001$), and coach teaching a lesson together in teacher's classroom ($t(134) = 3.58$, $p < .001$). CFC teachers also reported greater emphasis on all content areas: planning and reflecting on instruction ($t(130) = 4.72$, $p < .001$), providing help during lesson enactment ($t(131) = 4.26$, $p < .001$), building theory underlying pedagogy ($t(132) = 3.99$, $p < .001$), and differentiating instruction ($t(132) = 4.19$, $p < .001$) (see Matsumura et al., 2012 for more detailed descriptions of the differences between CFC coaching and literacy coaching in the comparison schools).

4. Discussion

The difficulties many students have comprehending texts with a high level of understanding underscore the need for interventions that provide better learning opportunities for students. Commensurate with other experimental and quasi-experimental studies of literacy-coaching programs (Powell et al., 2010; Sailors & Price, 2010), our study supports CFC as a strategy for improving both the quality of teaching and reading achievement in schools serving high numbers of minority and ELL students from low-income families. Importantly, this study helps us understand how a literacy-coaching program promotes student reading achievement, directly and indirectly by presenting evidence supporting the theory of action underlying the CFC program.

How did the CFC program affect student reading achievement? As hypothesized, the CFC program had a positive effect on the quality of classroom text discussions (Hypothesis 1a). Teachers in the CFC schools participated in coaching that focused on increasing teachers' ability to plan and enact high-quality text discussions through the application of Questioning the Author (QtA) techniques (Beck & McKeown, 2006). While the CFC teachers did not participate in coaching at the levels intended by the program developers, they did receive significantly more instructional support from their coaches (e.g., support for planning and reflecting on instruction and support in enacting lessons) than the teachers in the comparison schools. By the end of the second year of the program, text discussions in the CFC schools were more interactive and rigorous than in the comparison schools.

Also as hypothesized, text discussion quality functioned as a key mechanism through which CFC achieved its effect on students' reading skills (Hypothesis 1b). Teachers supported students to develop a deeper understanding of a text by posing high-level questions and pressing students to support their claims in discussions. Their students were more likely to build on each other's contributions, provide text-based evidence to support their claims, and discuss the high-level questions in support of developing a deeper understanding of a text, and, by the end of that academic year, students in the CFC schools demonstrated significantly higher reading achievement than their comparison group peers. These findings are aligned with research identifying high-level text discussions as a key instructional strategy for increasing students' reading comprehension skills (Murphy et al., 2009; Saunders & Goldenberg, 2007).

Our multi-level mediation model also allowed us to explore the moderating effect of students' language status. Frequently, language minority students are educated alongside native speakers of their country's language (e.g., Droop & Verhoeven, 1998; Stuart, 1999). School-level professional development programs are needed that increase teaching quality, learning opportunities, and the academic achievement of linguistically diverse populations of learners. The direct effect of the CFC program helped close the gap in reading achievement between ELL and non-ELL students in our study. The mediating effect of classroom discussion quality, as measured in our design, was similar for ELLs and non-ELLs, a finding consistent with other research linking higher-quality text discussions to improved reading achievement for both populations of students (e.g., Murphy et al., 2009; Saunders & Goldenberg, 2007). Since urban schools in the United States are likely to serve increasing numbers of ELL students (Kindler, 2002), this finding has important implications for the future.

Why did CFC appear to be especially effective for ELLs? Higher quality discussions and more frequent text discussions may expose students in general to a greater range of academic vocabulary and provide them with the opportunity to express themselves using new vocabulary. Vocabulary development may be especially important

for ELL students since it is critical to the development of reading comprehension skills (Lesaux, Kieffer, Faller, & Kelley, 2010; Snow, Lawrence, & White, 2009). A recent study of a vocabulary intervention in urban middle schools showed stronger effects for ELL students compared to their English-only peers (Snow et al., 2009). Increased vocabulary exposure could explain why the gap between the ELL students and the non-ELL peers closed in the CFC schools. Further research is needed to understand the CFC effect on ELL students' reading skills, and to investigate whether the pattern of effects we identified in our study would generalize to populations of low-income language-minority students outside of the United States.

Our finding that discussion quality mediated 18 percent of the total moderated effect on student achievement raises questions about other classroom factors that may play a role in student reading achievement. Possibly, how we measured discussion quality – the mediator in our analysis – did not capture the full range of instructional behaviors or classroom processes that contributed to student learning. For example, instructional behaviors that explicitly build students' academic vocabulary are not measured by the IQA (Stahl, Hare, Sinatra, & Gregory, 1991). Another possible unmeasured classroom-level mediator is the amount of classroom time teachers devoted to high-quality text discussions over the academic year (studies described in Murphy et al., 2009). The CFC program may have increased the frequency and length of text discussions at the same time that it raised the quality of those discussions seen in the classroom. Increasing the amount of time students are exposed to high-quality text discussions would increase students' opportunities to practice more complex and abstract forms of language, learn more vocabulary, and apply critical thinking skills – activities that are associated with students' reading achievement (Goldenberg, 2010; Nystrand, 2006; Stahl et al., 1991).

Other mechanisms at the school and district level may account for a portion of CFC's effectiveness. As a complex, multi-level intervention based on the idea that substantial changes in student learning require learning opportunities for actors at every level of the school system (Resnick & Spillane, 2006), CFC provides professional development to district leaders, principals, and coaches, as well as teachers. We targeted discussion quality as the mediator in our analysis based on the assumption that the key causal agents of an educational intervention's effect on student outcomes are located in instruction (Cohen et al., 2003). Decisions and activities at other levels of the school system, such as principal engagement in teachers' professional development and teachers' work with coaches in grade-level teams, could function as additional mediators of the CFC program's effect on student learning (McLaughlin & Talbert, 2006; Quint, Akey, Rappaport, & Willner, 2007). In order to improve the design and maximize the delivery of educational interventions (Pituch et al., 2010) further research is needed to understand the multiple mechanisms through which complex and multi-level interventions like CFC impact student learning. Our study underscores the difficulty and complexity of studying the effects of reform models that target both teaching processes and the organizational conditions in schools and districts that facilitate instructional improvement.

4.1. Limitations

Our study is limited by multiple factors that include instrumentation, attrition, response bias, and less-than-intended program implementation. We discuss these limitations in the following sections.

4.1.1. Instrumentation

More precise measures of the frequency of text discussions and vocabulary instruction, as well as the process of making meaning through those discussions, might better explain how CFC achieved its desired effects on student outcomes. In addition to teacher surveys, collecting data from sources such as teacher and coach logs and observation of coaching sessions would have strengthened our study by providing corroborating evidence of teachers' participation, and richer information about the school-level implementation of the CFC program. Notably, however, valid measures of coaching quality did not exist at the time our data were collected.

4.1.2. Attrition and response bias

Districts serving large numbers of low-income and minority students often experience high rates of teacher mobility (Henusheck et al., 2004). Nearly half of the teachers in our study left their school by the end of the first year of the trial. Although teacher attrition rates did not differ across conditions and were unrelated to demographic characteristics and student achievement, we acknowledge that the teachers who left their schools may have differed in unmeasured ways from the teachers who remained.

4.1.3. Implementation

Teachers reported participating in coaching activities in the second trial year at a lower-than-intended level. What effect the CFC program might have had on teaching and learning if implemented at the intended level remains an unanswered question. Detecting significant program effects in spite of this limitation suggests that less coaching than expected by the CFC program developers may be sufficient to achieve a significant benefit. How much coaching is necessary to achieve desired outcomes? To our knowledge, this question has not been investigated in an experimental study, and our project was not designed to provide an answer. The amount of coaching necessary to achieve target goals likely would vary as a function of the coaching intervention (e.g., how coaches work with teachers and what they focus on), the alignment of the assessed outcomes to the coaching intervention, and teachers' levels of skill and commitment to the intervention's goals. Examining the complex dynamic between these factors is a prime area for future research that could provide directions for optimizing the use of coaching resources.

4.2. Conclusion

The CFC study is one of the very few that has applied an experimental design to assess the effectiveness of school-based literacy coaching programs on both teaching and reading outcomes, and the only study to our knowledge that has used mediation analysis to examine the process by which a school-based coaching program achieves its effects. This study contributes to our understanding of the way a literacy coaching program can impact student achievement as well as adding to the body of evidence supporting a carefully designed literacy-coaching program as an effective strategy for increasing both teaching and learning outcomes (Biancarosa et al., 2010; Neuman & Cunningham, 2009; Powell et al., 2010; Sailors & Price, 2010). The application and analyses of our three-level mediation model illustrate how questions about multi-level interventions in educational field experiments involving mediation and moderation of treatment effects can be answered (Pituch et al., 2010).

Additional research is needed to examine the effectiveness and feasibility of adopting CFC on a widespread scale. Other research shows that coaching programs can be very difficult to implement with fidelity on a broad scale (e.g., Marsh et al., 2008), and CFC in specific requires a significant amount of resources to implement. Education authorities (e.g., school districts) that adopted CFC would need to pay both coaches and a person (likely an external provider) to provide professional development to coaches and leaders, in addition to allowing coaches and principals to be away from their buildings for multiple days. Questions remain regarding whether education authorities would be willing, or able, to allocate these kinds of resources to implement the program as intended.

Table A.1
Instructional Quality Assessment (IQA) ratings of quality of classroom text discussions.

Dimension of classroom text discussions	Rating codes			
	4	3	2	1
Widespread student participation in the discussion	At least 75% of the students participated throughout the discussion.	50–74% of the students participated in the discussion.	25–49% of the students participated in the discussion.	Less than 25% of the students participated in the discussion.
Teacher connects students' contributions to each other	The teacher consistently connects speakers' contributions to each other and shows how ideas/positions shared during the discussion relate to each other (e.g., "What I hear you saying is that the character has changed from the beginning to the end of the book which is similar to Alex's idea that the character has matured throughout the book.").	Twice during the lesson the teacher connect speakers' contributions to each other and shows how ideas/positions relate to each other.	At one or more points during the discussion, the teacher restates speakers' contributions, but does not show how ideas/positions relate to each other. No follow-up questions are asked after speakers' contributions (e.g., S1 – "The Sheriff is brave." T – "Ok you think the Sheriff is brave. Good. Robert?" S2 – "He's nice. " T = "You're saying he's nice. Ok. Anyone else?") OR only one strong effort is made to connect speakers' contributions to each other.	Teacher does not make any effort to link or restate speakers' contributions.
Students build on each other's contributions	Students consistently connect their contributions to each other and show how ideas/positions shared during the discussion relate to each other (e.g., "I agree with Alex because…"). The connections focus on and are relevant to the lesson content.	Twice during the lesson the students connect their contributions to each other and show how ideas/positions shared during the discussion relate to each other.	Students connect their contributions to each other, but do not show how their ideas/positions relate to each other (e.g., "I disagree with Alex.") OR only one strong effort is made by a student to connect their contributions to each other.	Students do not connect their contributions to the ideas expressed by other students.
Teacher presses students to support their claims	The teacher consistently asks students to explain their reasoning or provide evidence for their claims (e.g., "Where in the text does it say that?"; "Tell me more about why you agree with Alex that [the character] has matured?").	Twice during the lesson the teacher asks students to explain their reasoning and/or provide evidence for their claims	The teacher makes one or more superficial or formulaic efforts to ask students to explain their reasoning or provide evidence for their claims OR only one strong effort is made to press students to explain their reasoning or support their claims with evidence.	The teacher does not ask students to explain their thinking or to provide evidence for their claims.
Students support their claims	Students consistently provide accurate and appropriate evidence for their claims, including frequent references to the text or prior classroom experience OR students explain their thinking, using reasoning in ways appropriate to the discipline (e.g., "I think that [the character] will be different because it says [in the story] that after he and Lacey started to be friends other kids started saying hi to him.").	Twice during the lesson students provide accurate and appropriate evidence for their claims, including frequent references to the text or prior classroom experience, OR students explain their thinking, using reasoning in ways appropriate to the discipline.	What little evidence is offered to back up claims is inaccurate, incomplete, or vague (e.g., "I think that he will make more friends now.") OR students make only one strong effort to provide evidence for their claims or explain their thinking.	Students do not back up their claims, OR students do not explain the reasoning behind their claims.
Academic rigor of the text used in the discussions	n/a	The text contains substantial "grist" for students to grapple within a group discussion. This grist is seen in the complexity of the content (theme, relationships between characters, etc.) and in the writer's craft (literary language, rich vocabulary, organizational structures). Illustrative texts include, "Esperanza Rising" by Pam Munoz Ryan (2000) and "Charlotte's Web" by E.B. White (1952).	The text contains some "grist" for students to grapple with during group discussion. There may be some degree of complexity in the content (theme, relationships between characters, etc.) and in the writer's craft (literary language, rich vocabulary, organizational structures). Illustrative texts include most excerpts from basal readers and short articles from children's magazines.	There is minimal "grist" for students to discuss to make meaning of the story. It may contain a simple narrative or basic information. The text may be a simplified version of a complex text, or a short excerpt from a workbook. Illustrative texts include most decodable texts, very short selections from children's magazines, and texts (such as picture books) that are intended for much younger students.
Academic rigor of the discussion about the text	Two or more times during the discussion the teacher asks students questions that focus on the underlying meanings or literary characteristics of a text, or requires students to apply critical thinking skills to comprehend the text (e.g., "What do you think she means when she says that she always thinks of her dad as 'the preacher'? What might this say about their relationship?")	Once during the discussion the teacher asks students a question that focuses on the underlying meanings or literary characteristics of a text, or requires students to apply critical thinking skills to comprehend the text	The teacher asks questions that guide students to construct a coherent, but surface-level summary of the text (e.g., "What does her father do when she brings the dog home?")	The teacher asks questions that guide students to recall fragmented, isolated facts from a text (i.e., the sequence of the questions to not build toward creating a coherent representation of the text), OR the teacher only guides students to discuss a topic that does not directly reference information from the text (e.g., "How many of you have a dog?" "How did you get your dog?").

Table A.2
Percent of teachers participating at different levels in literacy coaching, Spring 2008.

Literacy coaching activity	CFC					Comparison				
	Never	1–3 times	4–6 times	Monthly	Weekly	Never	1–3 times	4–6 times	Monthly	Weekly
	%	%	%	%	%	%	%	%	%	%
Coach met with me and other teachers in grade-level meetings	1	11	15	36	38	4	13	20	31	31
Coach met with me individually	2	30	25	19	24	23	30	16	20	11
Coach observed me teaching 30 or more minutes	10	44	20	20	8	59	21	11	5	5
Coach taught a model lesson for me	10	53	17	12	9	64	30	5	0	2
Coach taught a lesson with me in my classroom	32	44	11	10	3	80	9	7	0	5

Note. Spring 2008 survey reports were provided by 93 CFC and 45 comparison teachers.

Table A.3
Percent of teachers reporting emphasis of focus in literacy coaching activities, Spring 2008.

Focus of activity	CFC				Comparison			
	None	Minor	Moderate	Major	None	Minor	Moderate	Major
	%	%	%	%	%	%	%	%
Emphasis on planning and reflecting on instruction	9	24	39	28	25	41	27	7
Emphasis on help during lesson enactment	5	29	38	28	23	44	26	7
Emphasis on building theory underlying pedagogy	19	33	28	20	55	25	11	9
Emphasis on differentiating instruction	26	18	37	19	36	39	23	2

Note. Spring 2008 survey reports were provided by 93 CFC and 45 comparison teachers.

Acknowledgments

The work described in the current article was supported by a research 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.

References

Applebee, A., Langer, J., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: classroom instruction and student performance in middle and high school English. *American Education Research Journal, 40*, 685–730. http://dx.doi.org/10.3102/00028312040003685.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research and Practice, 20*(1), 50–57. http://dx.doi.org/10.1111/j.1540-5826.2005.00120.x.

Beck, I. L., & McKeown, M. B. (2001). Inviting students into the pursuit of meaning. *Educational Psychology Review, 13*(3), 225–241. http://dx.doi.org/10.1023/A:1016671722022.

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 Inc.

Beck, I. L., McKeown, M. G., Sandora, C., Kucan, L., & Worthy, J. (1996). Questioning the author: a yearlong classroom implementation to engage students with text. *Elementary School Journal, 96*(4), 385–414. http://dx.doi.org/10.1086/461835.

Biancarosa, G., Bryk, A., & Dexter, E. (2010). Assessing the value-added effects of literacy collaborative professional development on student learning. *Elementary School Journal, 111*(1), 7–34. http://dx.doi.org/10.1086/653468.

Blarney, K. L., Meyer, C. K., & Walpole, S. (2009). Middle and high school literacy coaches: a national survey. *Journal of Adolescent and Adult Literacy, 52*(4), 310–323. http://dx.doi.org/10.1598/JAAL.52.4.4.

Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In H. S. Bloom (Ed.), *Learning more from social experiments: Evolving analytic approaches* (pp. 115–172). New York: Russell Sage Foundation.

Boruch, R. F., & Foley, E. (2000). The honestly experimental society. In L. Bickman (Ed.), *Validity and social experiments: Donald Campbell's legacy* (pp. 193–239). Thousand Oaks, CA: Sage, (No DOI found).

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press (No DOI found).

Buly, M. R., & Valencia, S. W. (2002). Below the bar: profiles of students who fail state reading assessments. *Educational Evaluation and Policy, 24*(3), 219–239. http://dx.doi.org/10.3102/01623737024003219.

Cazden, C. (2001). *Classroom discourse*. Portsmouth, NH: Heinemann.

Chall, J. S., Jacobs, V. A., & Baldwin, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press (No DOI found).

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction and research. *Educational Evaluation and Policy Analysis, 35*(2), 119–142. http://dx.doi.org/10.3102/01623737025002119.

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

Costa, A. L., & Garmston, R. J. (2002). *Cognitive coaching: A foundation for renaissance schools* (2nd ed.). Norwood, MA: Christopher-Gordon Publisher.

Droop, M., & Verhoeven, L. (1998). Background knowledge, linguistic complexity, and second-language reading comprehension. *Journal of Literacy Research, 30*, 253–271. http://dx.doi.org/10.1080/10862969809547998.

Duessen, T., Coskie, T., Robinson, L., & Autio, E. (2007). *"Coach" can mean many things: Five categories of literacy coaches in reading first*. REL Northwest (REL 2007-No. 005). IES: US Department of Education.

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

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. (1993). Instructional conversations: promoting reading comprehension through discussion. *The Reading Teacher, 46*(4), 316–326.

Goldenberg, C. (2010). Reading instruction for English language learners. In M. Kamil, P. D. Pearson, E. Moje, & P. Afflerbach (Eds.). *Handbook of reading research, Vol. IV*. Newark, DE: International Reading Association.

Hawley, W. D., & Valli, L. (1999). The essentials of effective professional development: a new consensus. In G. Sykes, & L. Darling-Hammond (Eds.), *Handbook of teaching and policy*. New York: Teachers College.

Henusheck, E. A., Kain, J. F., & Rivkin, S. (2004). Why public schools lose teachers. *The Journal of Human Resources, 39*(2), 326–354. http://dx.doi.org/10.2307/3559017.

Joyce, B., & Showers, B. (1988). *Student achievement through staff development*. New York: Longman.

Kelley, J. G., Lesaux, N. K., Kieffer, M. J., & Faller, S. E. (2010). Effective academic vocabulary instruction in the urban middle school. *Reading Teacher, 64*(1), 5–14. http://dx.doi.org/10.1598/RT.64.1.1.

Kieffer, M. J. (2008). Catching up or falling behind? Initial English proficiency, concentrated poverty, and the reading growth of language minority learners in the United States. *Journal of Educational Psychology, 100*(4), 851–868. http://dx.doi.org/10.1037/0022-0663.100.4.851.

Kindler, A. L. (2002). *Survey of the states' limited English proficient students and available educational programs and services, 2000–2001 summary report*. Washington, DC: National Clearinghouse for English Language Acquisition and Language Instruction Educational Programs.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363–394. http://dx.doi.org/10.1037/0033-295X.85.5.363.

Lee, J., Grigg, W., & Donahue, P. (2007). *The nation's report card: Reading 2007 (NCES 2007-496)*. Washington, D.C: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for

linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45*(2), 196–228. http://dx.doi.org/10.1598/RRQ.45.2.3.

Lesaux, N. K., Rupp, A. A., & Siegel, L. S. (2007). Growth in reading skills of children from diverse linguistic backgrounds: findings from a 5-year longitudinal study. *Journal of Educational Psychology, 99*(4), 821–834. http://dx.doi.org/10.1037/0022-0663.99.4.821.

Lovett, M. W., Lacerenza, L., De Palma, M., Benson, N. J., Steinbach, J. A., & Frijters, J. C. (2008). Preparing teachers to remediate reading difficulties in high schools. What is needed for effective professional development? *Teaching and Teacher Education, 24*(4), 1083–1097. http://dx.doi.org/10.1016/j.tate.2007.10.005.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593–614. http://dx.doi.org/10.1146/annurev.psych.58.110405.085542.

MacKinnon, D. P., Fritz, M. S., Williams, J., & Lockwood, C. M. (2007). Distribution of the product confidence limits for the indirect effect: program PRODCLIN. *Behavior Research Methods, 39*, 384–389. http://dx.doi.org/10.3758/BF03193007.

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.

Matsumura, L. C., Garnier, H. E., & Resnick, L. (2010). Implementing literacy coaching: the role of school social resources. *Educational Evaluation and Policy Analysis, 3*(1), 35–62. http://dx.doi.org/10.3102/0162373710363743.

Matsumura, L. C., Garnier, H. E., Slater, S. C., & Boston, M. (2008). Toward measuring instructional interactions 'at scale.'. *Educational Assessment, 13*(4), 267–300. http://dx.doi.org/10.1080/10627190802602541.

Matsumura, L. C., Garnier, H. E., & Spybrook, J. (2012). The effect of content-focused coaching on the quality of classroom text discussions. *Journal of Teacher Education, 63*(3), 214–228. http://dx.doi.org/10.1177/0022487111434985.

McKeown, M. G., Beck, I. L., & Blake, R. G. K (2009). Rethinking reading comprehension instruction: a comparison of instruction for strategies and content approaches. *Reading Research Quarterly, 44*(3), 218–253. http://dx.doi.org/10.1598/RRQ.44.3.1.

McLaughlin, M. W., & Talbert, J. E. (2006). *Building school-based teacher learning communities: Professional strategies to improve student achievement*. New York: NY: Teachers College.

Michaels, S., O'Connor, C., & Resnick, L. B. (2008). Deliberative discourse idealized and realized: accountable talk in the classroom and in civic life. *Studies in Philosophy and Education, 27*(4), 283–297. http://dx.doi.org/10.1007/s11217-007-9071-1.

Murphy, P. K., Wilkinson, I. A. G., Soter, A.,O., Hennessey, M. N., & Alexander, J.,F. (2009). Examining the effects of classroom discussion on students' comprehension of text: a meta-analysis. *Journal of Educational Psychology, 101*, 740–764. http://dx.doi.org/10.1037/a0015576.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*(2), 532–566. http://dx.doi.org/10.3102/0002831208328088.

Nystrand, M. (2006). Research on the role of classroom discourse as it affects reading comprehension. *Research in the Teaching of English, 40*, 392–412.

Pituch, K. A., Murphy, D. L., & Tate, R. L. (2010). Three-level models for indirect effects in school- and class-randomized experiments in education. *Journal of Experimental Education, 78*, 60–95. http://dx.doi.org/10.1080/00220970903224685.

Powell, D. R., Diamond, K. E., Burchinal, M. R., & Koehler, M. J. (2010). Effects of an early literacy professional development intervention on head start teachers and children. *Journal of Educational Psychology, 102*(2), 299–312. http://dx.doi.org/10.1037/a0017763.

Quint, J. C., Akey, T. M., Rappaport, S., & Willner, C. J. (2007). *Instructional leadership, teaching quality, and student achievement: Suggestive evidence from three urban school districts*. MDRC, Retrieved on 14.10.08 from. http://www.mdrc.org/publications/470/html.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*(2), 173–185. http://dx.doi.org/10.1037/1082-989X.2.2.173.

Raudenbush, S. W. (March 3, 2011). *Modeling mediation: Causes, markers and mechanisms*. Washington, DC: Address at the Society for Research on Educational Effectiveness.

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

Raudenbush, S. W., Martinez, A., & Spybrook, J. (2007). Strategies for improving precision in group-randomized experiments. *Educational Evaluation and Policy Analysis, 29*(1), 5–29. http://dx.doi.org/10.3102/0162373707299460.

Resnick, L. B., & Spillane, J. P. (2006). From individual learning to organizational designs for learning. In L. Verschaffel, F. Dochy, M. Boekaerts, & S. Vosniadou (Eds.), *Instructional psychology: Past, present and future trends: Sixteen essays in honour of Erik De Corte (advances in learning and instruction series)*. Oxford: Pergamon.

Sailors, M., & Price, L. (2010). Professional development for cognitive reading strategy instruction. *Elementary School Journal, 110*, 301–323. http://dx.doi.org/10.1086/648980.

Saunders, W., & Goldenberg, C. (2007). The effects of an instructional conversation on English Language Learners' concepts of friendship and story comprehension. In R. Horowitz (Ed.), *Talking texts: How speech and writing interact in school learning* (pp. 221–252). Mahwah, NJ: Erlbaum.

Snow, C. E., Lawrence, J. F., & White, C. (2009). Generating knowledge of academic language among urban middle school students. *Journal of Research on Educational Effectiveness, 2*(4), 325–344. http://dx.doi.org/10.1080/19345740903167042.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). Washington, DC: American Sociological Association.

Stahl, S. A., Hare, V. C., Sinatra, R., & Gregory, J. F. (1991). Defining the role of prior knowledge and vocabulary in reading comprehension: the retiring of number 41. *Journal of Literacy Research, 23*, 487–508. http://dx.doi.org/10.1080/10862969109547755.

Stanulis, R. N., Little, S., & Wibbens, E. (2012). Intensive mentoring that leads to change in beginning elementary teachers' learning to lead classroom discussions. *Teaching and Teacher Education, 28*, 32–43. http://dx.doi.org/10.1016/j.tate.2011.08.007.

Staub, R. (2004). Transforming educational theory into usable knowledge: a case of co-constructing tools for lesson design and reflection. In B. Ralle, & I. Eilks (Eds.), *Quality in practice-oriented research in science education* (pp. 41–52). Aachen, Germany: Shaker.

Staub, F. C., West, L., & Bickel, D. D. (2003). What is content-focused coaching? In L. West, & F. C. Staub (Eds.), *Content-focused coaching. Transforming mathematics lessons* (pp. 1–17) Portsmouth, NH: Heinemann.

Stigler, M. H., Kugler, K. C., Komro, K. A., Leshabari, M. T., & Klepp, K. I. (2006). AIDS education for Tanzanian youth: a mediation analysis. *Health Education Research: Theory and Practice, 21*(4), 441–451. http://dx.doi.org/10.1093/her/cyh070.

Stuart, M. (1999). Getting ready for reading: early phoneme awareness and phonics training improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology, 69*(4), 587–605. http://dx.doi.org/10.1348/000709999157914.

Technical Digest 2007-2008 (n.d.). *Texas education agency*. Retrieved from http://www.tea.state.tx.us/student.assessment/techdigest/yr0708/.

Teemant, A., Wink, J., & Tyra, S. (2011). Effects of coaching on teacher use of sociocultural instructional practices. *Teaching and Teacher Education, 27*(4), 683–693. http://dx.doi.org/10.1016/j.tate.2005.03.002.

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

Van Keer, H., & Verhaeghe, J. P. (2005). Comparing two teacher development programs for innovating reading comprehension instruction with regard to teachers' experiences and student outcomes. *Teaching and Teacher Education, 21*, 543–562. http://dx.doi.org/10.1016/j.tate.2005.03.002.

VanderWeele, T. J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology, 20*, 18–26. http://dx.doi.org/10.1097/EDE.0b013e31818f69ce.

Veenman, S., & Denessen, E. (2001). The coaching of teachers: results of five training studies. *Educational Research and Evaluation, 7*(4), 385–417. http://dx.doi.org/10.1076/edre.7.4.385.8936.

Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT.

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

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, TX: National Staff Development Council.

What Works Clearinghouse. (2008). *WWC procedures and version 2 standards handbook*. Retrieved 02.12.09 from. http://ies.ed.gov/ncee/wwc/references/idocviewer/doc.aspx?docid=19&tocid=1.