The impact of inquiry-based professional development on teachers' capacity to integrate literacy instruction in secondary subject areas

Cynthia Greenleaf*, Cindy Litman, Stacy Marple

Strategic Literacy Initiative, WestEd, United States

HIGHLIGHTS

- PD is critical for building teachers' capacity to integrate literacy and content.
- PD lessons showed greater integration of complex literacy practices and content.
- PD lessons showed less delivery of content and more close reading and use of text.
- Reading and reasoning in PD classrooms served disciplinary knowledge building.
- Inquiry may be the key behind-the-scenes player in effective PD.

ABSTRACT

This mixed-methods study explores the impact of an inquiry-based PD model designed to support teachers in integrating literacy and content learning. The study drew on classroom observations of 34 ELA, history, and science teachers with reputations for good instructional practice. An analysis of video data from observed lessons revealed that a subset of 16 teachers who had previously participated in Reading Apprenticeship PD offered significantly more opportunities and support for complex literacies and greater integration of literacy and content learning. Findings have implications for PD with the potential to improve literacy instruction in secondary content area classrooms.
standards and assessments, curriculum materials, and teaching practices (Opfer, Kaufman, & Thompson, 2016; Porter, McMaken, Hwang, & Yang, 2011).

There is general agreement that professional development (PD) must be a partner to curricular reform in building teachers’ capacity for teaching literacy alongside content (Darling-Hammond, Hyler, & Gardner, 2017; Heller & Greenleaf, 2007; Kaufman et al., 2016; Kober & Rentner, 2012). However, despite progress in identifying general features of high quality PD (Desimone, 2009), few studies have investigated PD that effectively supports teachers in integrating literacy and content learning, particularly among non-English language arts (ELA) teachers, who are now being asked to address ELA literacy standards in their instruction (Cobb & Jackson, 2011; Kaufman et al., 2016; Wilson, 2013). In addition, although recent models of PD view teacher change as an ongoing process (Clarke & Hollingsworth, 2002; Tabak & Radinsky, 2015), few studies have examined the impact of specific PD programs on teachers’ practices beyond the PD year (Coburn, 2003; Kennedy, 2016). Hargreaves and Goodson (2006) suggest that the large body of research based on early implementation may provide an overly optimistic picture of the impact of PD. Indeed, studies that have followed teachers beyond the PD year confirm that positive effects of PD often fade over time (Coburn, 2003; Hargreaves & Goodson, 2006). However, this is not always the case. Consistent with a view of teachers as “active learners shaping their professional growth through reflective participation in professional development programs and in practice” (Clarke & Hollingsworth, 2002, p. 948), a small number of studies indicate that positive impacts of PD can potentially increase with time (Franke, Carpenter, & Levi, 2001; Glazerman et al., 2010; Harris & Sass, 2011; Heller, Daehler, Wong, Shinohara, & Miratrix, 2012; Kennedy, 2010; Wang, Hsu, Reeves, & Coster, 2014). Based on evidence of a lag effect, Kennedy (2016) argues that “an important and under-emphasized question in research on PD is whether PD produces enduring changes in practice rather than temporary compliance” (p. 7).

The study reported here represents a contribution to this research agenda by examining the impact of a PD model designed to develop teachers’ capacity to integrate literacy and content learning. The study draws on a larger program of design-based research that invited teachers to play a central role in the design of instructional approaches to support disciplinary argumentation from multiple sources. To inform the design work, we carried out classroom observations of 34 middle and high school ELA, history, and science teachers prior to designing any instructional intervention. All the teachers had reputations for good instructional practice. However, somewhat different recruitment and selection processes at the two research sites resulted in two subgroups of teachers: a group of 16 teachers who had previously participated in Reading Apprenticeship PD and a group of 18 teachers who had not. Although we expected to see promising practices in all these classrooms, an initial qualitative analysis of field notes and lesson artifacts revealed different patterns of practice in the two groups, with teachers who had participated in Reading Apprenticeship PD offering more opportunities and support for the kinds of learning envisioned by current literacy reform efforts. Of particular interest, many teachers who had participated in Reading Apprenticeship PD had done so several years prior to the observational study reported here, suggesting enduring effects of Reading Apprenticeship PD on teacher practice and thereby on students’ opportunity to learn (OTL). To more systematically investigate the impact of the Reading Apprenticeship PD model, utilizing a mixed methods approach, we conducted an analysis of literacy OTL offered by teachers in PD and comparison classrooms based on video data from observed teachers. Here we report findings from this analysis.

1. Conceptual framework

The conceptual framework guiding this study is based on opportunity to learn, the notion that educational outcomes depend on educational opportunity (Porter, 1995). Classroom OTL, operationalized as learning opportunities presented by the teacher, has strong effects on student learning, even after controlling for a variety of teacher and student variables (Abedi & Herman, 2010; Boscardin et al., 2005; Heafner & Fitchett, 2015; Wang, Haertel, & Walberg, 1997). The theory of action underlying the study posits that Reading Apprenticeship PD experienced by teachers in the study enabled them to provide opportunities and support for students to carry out close, intellectually engaged reading and meaning-making about text, ultimately leading to increased student academic engagement and achievement.

2. Research questions

The overarching question guiding the study was: Did teachers’ participation in Reading Apprenticeship PD have an impact on the literacy learning opportunities they offered students? Specifically, we compared “PD” and comparison teachers on instructional practices explicitly emphasized by Reading Apprenticeship PD, as well as those called for with the advent of new standards and policies:

- **Centrality of text.** Reading Apprenticeship PD emphasizes the centrality of texts to the authentic intellectual work of learning. This is also an explicit goal of the Common Core. Therefore, we asked:
  - What opportunities did teachers provide for students to work with text, in contrast to learning content through teacher lecture or explanation?

- **Focus on complex literacies.** The PD provided to teachers in the current study focused on building students’ capacities to carry out close, intellectually engaged reading and meaning-making about text. Because it occurred prior to the onset of recent literacy reforms, it did not explicitly address other complex literacy practices emphasized by new standards, specifically argumentation and multi-text literacy. To investigate the extent to which participation in Reading Apprenticeship PD led to changes in understanding and practice that prepared teachers to respond to the new policy context as well as to sustain instructional practices explicitly targeted by Reading Apprenticeship PD, we asked:
  - What opportunities did teachers provide for students to engage in complex literacy practices?

- **Integration of literacy and content learning.** The perception that literacy instruction competes with content learning is a significant barrier to literacy OTL across subject areas (Misulis, 2009; Ness, 2008; Reed, 2009). To examine teachers’ capacity to integrate literacy and content learning, we asked:
  - How were literacy and content learning related to one another?

- **Shifting the intellectual work of learning to students.** Finally, Reading Apprenticeship PD provides tools and approaches for developing students’ independent engagement in the intellectual work of learning, also an explicit focus of Common Core aligned standards. Using grouping structure as an indicator, we looked at opportunities for students to engage in various learning interactions and supports, particularly opportunities for peer-directed work that potentially shifts interpretive authority to students. We asked:
• How did teachers allocate time to various kinds of learning interactions and supports for literacy learning?
• What opportunities did teachers provide for students to engage in peer-directed groups focused on making meaning from text?

3. Literature review

3.1. Challenges to the impact of professional development

Research paints an inconsistent and disappointing picture of the long-term impact of PD on teachers’ literacy instruction and students’ literacy achievement. For example, a large-scale study of public school teachers and students over a four-year period found no long-term effects of PD on reading instruction, either during the PD year or in subsequent years (Jacob & Lefgren, 2004). These findings have prompted many researchers to look at other factors that influence the implementation and sustainability of PD.

Prior knowledge, beliefs, and experience. Research has shown that the knowledge, beliefs, and experiences teachers bring to PD exert a powerful influence on their development of new concepts and understandings (Minor, Desimone, Caines Lee, & Hochberg, 2016; Spillane, Reiser, & Reimer, 2002). Rather than implementing new approaches as intended, Huberman (1995) notes that teachers may engage in a process of “bricolage” or tinkering by assimilating new ideas into current practices and adjusting them to fit prior knowledge and experiences. Research suggests that many teachers have responded in this way to PD aimed at improving literacy instruction. Despite the wide gulf between current literacy practices and instruction mandated by literacy reforms (Kober & Rentner, 2012; Opfer et al., 2016; Porter et al., 2011), two-thirds of teachers at the onset of the Common Core anticipated making only slight adjustments in their teaching to align with the standards (ACT, Inc., 2013b). Although more than three-quarters of teachers have subsequently participated in PD aimed at implementing state literacy standards (Kaufman et al., 2016), there has been little substantive change in literacy instruction (Santelises & Dobrowksi, 2015). Indicative of bricolage, a recent study found that the overwhelming majority of ELA, science, and social studies lessons ostensibly aligned with the Common Core exhibited “window dressing” that missed the deeper intent of the instructional shifts mandated by the standards (Santelises & Dobrowski, 2015, p. 7).

Contextual factors. Research also shows that teaching context plays a major role in whether PD is implemented and sustained (Johnson, 2009; Kraft & Papay, 2014; Little, 2006; McLaughlin & Mitra, 2001). A synthesis of research on PD aimed at improving content area literacy instruction found that teachers with more district- and school-level support implemented PD strategies with greater fidelity, even when they initially believed the PD was not appropriate for their students (Reed, 2009). Conversely, even when PD provides teachers with knowledge and skills to integrate literacy and content instruction, workplace conditions can potentially undermine PD’s impact (Kaufman et al., 2016; Opfer et al., 2016). A longitudinal study of a PD intervention designed to increase language proficiency for English learners provides an illustrative case (Bos et al., 2012). During the three-year study, three participating schools were “consolidated,” nearly 30% of students in the original cohort left the school, and staffing and logistical issues prevented most eligible teachers from participating in PD services as intended. Although the program itself instantiated core features of effective PD, it is perhaps not surprising that the study found no significant impact on teacher or student outcomes. Even under less extreme circumstances, the extent to which schools allocate sufficient time, planning, and resources exerts a major influence on the impact of PD (Adelman & Taylor, 2003; Coburn, Russell, Kaufman, & Stein, 2012; Han & Weiss, 2005; Hargreaves & Goodson, 2006; Kraft & Papay, 2014; McLaughlin & Mitra, 2001). Reed (2009) found that lack of time to prepare and lack of access to materials were universal impediments to implementing and sustaining PD aimed at integrating literacy and content learning.

Characteristics of PD. Despite significant obstacles, a large body of research demonstrates that PD can nonetheless help teachers reconstruct their instructional practices (Wei, Darling-Hammond, & Adamson, 2010). In a widely cited study, Desimone (2009) argues that there is sufficient empirical evidence to establish a consensus on five core features of high quality PD: (a) focus on subject matter content and how students learn that content, (b) active learning opportunities, coherence with teachers’ existing practices and policy contexts, (c) sufficient duration, and (d) collective participation. Yet research indicates that PD based on this framework sometimes fails to produce expected outcomes (Opfer & Pedder, 2011). A recent evaluation brief from the Institute of Education Sciences reached the sobering conclusion that “the field does not yet fully understand how to ensure that teacher PD leads to measurable improvements in student learning” (Garet, Heppen, Wolters, Smith, & Yang, 2016, p. 11).

The challenge of re-enactment. Bringing a different lens to the question of what makes PD effective and enduring, Kennedy (2016) recently argued that teacher change requires not only re-understanding, but also re-enacting core teaching practices. Reanalyzing experimental studies of PD along this dimension, Kennedy found that effective PD not only focused on building new knowledge, but also actively supported re-enactment of teaching practices by intellectually engaging teachers with PD content and helping them translate new ideas and approaches into their own systems of practice. Furthermore, teacher and student outcomes from these effective programs tended to increase with time. Although overshadowed recently by a focus on features of effective PD codified by Desimone (2009), Kennedy’s findings highlight “the importance of giving teachers the time and opportunity to make their own sense of new ideas” (p. 27). Inquiry into the complexities of enacting new curricular and instructional approaches may be especially important in the context of educational reforms, when teachers are being asked to teach in ways that are substantially different from how they were taught or how they learned to teach (Berman & McLaughlin, 1978; Borko, Jacobs, & Koellner, 2010; Spillane et al., 2002).

3.2. Contribution of the current study

Long considered an essential element of PD focused on inquiry- and problem-based mathematics and science instruction (Loucks-Horsley, Stiles, & Hewson, 1996; Supovitz & Turner, 2000), inquiry has not been a salient feature of literacy PD (Reznitskaya & Hochberg, 2016). The current study extends this line of research by investigating the enduring impact of an inquiry-based PD model that engages teachers in making sense of the complexities and problematic aspects of integrating literacy and content area learning.

Before turning to methods and results, to make sense of the contrasts that emerged between the two groups of teachers, we briefly describe the Reading Apprenticeship instructional framework and PD model, as well as specific features of the 7-day institutes in which members of the PD group had previously participated. Elsewhere we have provided detailed descriptions of Reading Apprenticeship PD (Greenleaf & Schoenbach, 2004; Greenleaf et al., 2011; Schoenbach, Greenleaf, & Murphy, 2017).
4. Reading Apprenticeship professional development

Reading Apprenticeship is a model of academic literacy instruction designed to improve students’ literacy skills and academic achievement. Based on understandings of the close relationship between curriculum reform and PD (Heller & Greenleaf, 2007), Reading Apprenticeship components include an instructional framework and associated PD model for secondary and post-secondary teachers across the academic subject areas. Both the instructional framework and the PD model are products of extended collaborative design research processes, informed by sociocultural learning theory and research in language and literacy development (Greenleaf & Schoenbach, 2004). Although predating current literacy reform efforts, Reading Apprenticeship is aligned with principles of reform initiatives that specify the need for advanced literacy skills and understandings across the content areas, not just in ELA.

Guided by the instructional framework (see Schoenbach, Greenleaf, & Murphy, 2012), reading instruction is integrated into subject area teaching, rather than being an instructional add-on or additional curriculum. Teachers across the subject areas learn how to build students’ capacities to carry out complex, intellectually engaged reading, make meaning, acquire academic and disciplinary language, read independently, and set personal goals for literacy development. Through an “apprenticeship” process and ongoing metacognitive conversations, subject area teachers explicitly teach students the tacit reasoning processes, strategies, and discourse rules that shape successful readers’ and writers’ work in their disciplines.

4.1. PD model

Mirroring the instructional framework, Reading Apprenticeship PD has been developed to transform teachers’ understanding of their role in students’ literacy development and to build teachers’ capacity for re-enacting literacy instruction in the academic disciplines (Greenleaf & Schoenbach, 2004; Schoenbach et al., 2017). The PD is inquiry-based, subject area focused, and designed to address teachers’ conceptual understandings as well as practical implementation needs. The model reflects the understanding that for practice to become truly responsive to the needs and varied contexts of teachers’ work, teachers must become adaptive and generative in their use of specific practices (Ball & Cohen, 1999; Gillis, 2014; Kennedy, 2016).

The 7-day Reading Apprenticeship PD institute experienced by the PD group immerses teachers in learning through models of practice that its designers intend for them to create in their own classrooms, supporting what Kennedy (2016) calls “re-enactment.” Teachers participate in carefully designed inquiries to help them unlock their own disciplinary expertise in relation to literacy. Most importantly, they collaboratively investigate student work, case studies of student literacy learning, and videotaped classroom lessons designed to foster new expectations of what their own students can accomplish. In PD sessions, they enact classroom routines to build student engagement, support student collaboration, and foster authentic discussion and problem-solving around course texts, all with the goal of learning new ways to support students’ thinking and learning with academic materials.

Table 1 describes core Reading Apprenticeship PD inquiry routines. Although they serve distinct purposes, inquiries are bundled to serve multiple learning goals at once. Teachers never analyze student reading performance without first doing their own reading process analyses and trying out any specific learning task students experience with that text(s). Reading process analysis and text and task analysis thus almost always accompany analyzing student work. These inquiries in turn often lead to analyzing the instructional design of the PD experience as a model of literacy instruction, and then to their own instructional design and reflections on learning. Professional learning and instructional planning are often informed and supported by professional reading and text-based discussion. Bundled inquiries thus allocate significant time, opportunity, and support for applying the reading, thinking, and talking routines and tools that teachers experience in PD to their own classroom. We describe a typical bundled inquiry below.

**Typical PD inquiry.** Superman and Me, a half-day inquiry engaging teachers in building understanding of literacy and noticing and responding to student thinking, begins with a reading process analysis of Superman and Me (Alexie, 1997), the text that teachers will see students reading in a classroom video. For this inquiry, teachers use a written form of think aloud, Talking to the Text, that begins as an individual reading and writing activity, is debriefed with a partner, then moves to a whole group metacognitive conversation during which the facilitator elicits and charts teachers’ reading and thinking processes to create a collaborative reading strategies list. Teachers then watch the classroom video using another Reading Apprenticeship metacognitive reading routine, a double entry Evidence/Interpretation notetaker designed to focus teachers’ attention on evidence, rather than leaping directly to interpretation (Schoenbach et al., 2012). After several cycles of viewing, note-taking, and small group sharing, teachers discuss what they notice about student reading, thinking, and talking and what supports their work with the whole group, while the facilitator probes for additional evidence.

The PD inquiry then offers opportunity and support for applying routines and tools teachers experienced themselves and in the video to their own classrooms. Repeating the cycle of inquiry they experienced with the classroom case, teachers plan and practice their own think aloud using texts they brought to the PD session. Taking turns playing the role of teacher and student with two different partners, they focus on opportunities to probe for deeper thinking, one of the most challenging aspects of Reading Apprenticeship classrooms. Finally, using a planning and reflection tool to help them think about key elements of Reading Apprenticeship instruction (Schoenbach et al., 2012, p. 337), teachers plan a lesson using their own course texts. As teachers share their lessons with the group, the facilitator focuses discussion on ways the lessons support students in doing the intellectual work of learning. In the closing routine, teachers reflect and consolidate the day’s learning. Throughout the 7-day institute, teachers engage in numerous half- or full-day inquiry cycles designed to support them in both re-understanding and re-enacting literacy instruction.

Multiple studies of the impact of Reading Apprenticeship PD on teacher learning and classroom practice have affirmed the effectiveness of this model of teaching and teacher learning (Fancsali et al., 2015; Greenleaf et al., 2011; Somers et al., 2010). However, the current study is the first to look at the long-term impact of Reading Apprenticeship PD on teachers’ classroom practice.

5. Data and methods

The data set consisted of videotapes of 71 lessons taught by 34 teachers, grades 6–12, from 22 San Francisco Bay and Chicago area urban and suburban schools. Observations were conducted between October 2010 and June 2011. At that time, the Common Core had been released, but implementation was not yet expected. Since then, widespread PD aimed at supporting literacy standards has yet to produce desired changes in instruction (Santelises & Dobrowski, 2015), and advocates of literacy reform as well as teachers themselves continue to seek out PD with the potential to improve literacy instruction (Kaufman et al., 2016). Therefore, we
Professional reading and text-based discussion. Enacted as reading and discourse routines for the classroom, professional reading and text-based discussions serve the dual purpose of increasing teachers’ understandings of literacy practices and modeling Reading Apprenticeship practices promoted for students. This means 1) never assigning reading without making time for text-based discussion and 2) enacting participation routines that support each teacher to meaningfully participate in discussion, such as providing time in the session to read and/or review reading in order to prepare for discussion; providing an overarching question or prompt to elicit discussion; enacting protocols for discussion that ensure responsive dialogue and equitable distribution of airtime; and using pedagogies that support concept development.

Reading process analysis (RPA). Reading process analysis (RPA) is a central routine of Reading Apprenticeship PD. Based on a Think Aloud protocol, this metacognitive activity is designed to make teachers’ thinking visible as they read complex texts. RPA variations include verbal and written, cross-discipline and discipline-specific inquiries into ordinarily invisible reading processes. By noticing and sharing their reading and thinking processes in a community of colleagues, teachers gain an understanding of the processes involved in making sense of texts, the variety of resources and strategies they and others bring to reading tasks, and a language for talking about these mental processes. These metacognitive conversations about reading give teachers practice in the kinds of “in the moment” conversations that will prepare their students in reading and reasoning about texts. RPAs are documented, frequently through listing the strategies used by participating readers, in a Reading Strategies List, for further analysis and discussion.

Text and task analysis. In Reading Apprenticeship PD, teachers learn to analyze texts and tasks as a professional habit in order to anticipate challenges and plan ways to support students through these challenges—rather than protect them from challenge through pre-teaching or reducing text complexity. The text and task analysis routine engages teachers in identifying not only the reasoning processes they use to make sense of text, but also the text features, language, content, and disciplinary perspectives that texts present and that students might find challenging. The analysis of text and task demands follows reading process analysis, during which teachers reflect on their mental processes for making meaning of texts. Shifting their attention to the text(s) and to the task they are asked to complete with the text(s), teachers then use a graphic organizer to identify aspects of the text(s) that present schema demands in different categories: world knowledge, topic knowledge, language and vocabulary, and disciplinary perspectives/discourse.

Analyzing student work. Although it has multiple aims, most importantly, the analyzing student work routine is designed to build teachers’ insight into learning and the needs and abilities of learners. A specific focus of this routine is dispelling teachers’ beliefs that their students cannot do complex work and thinking. The routine is asset oriented—what are students doing well? It leads to instructional decision-making—what do students need to learn to do better? Tools and protocols are used that ground interpretations and conceptions about student work in evidence and promote evidence-based discussion, including an evidence/interpretation notetaker and discussion protocol. Similarly, rubrics and student learning goals support teachers in identifying strengths and instructional needs. In analyzing student work, teachers sometimes work with written work samples and sometimes from videotapes of student performances.

Artifact-grounded reflections on practice. Assignment analysis is a powerful tool for reflecting on classroom practices and student learning experiences. One routine for helping teachers make connections between activities in the professional development setting and classroom practices is asking teachers to bring in artifacts from a lesson reflect on their teaching and students’ learning, and discuss their successes and challenges with one another.

Analyzing the instructional design of Reading Apprenticeship PD. Reading Apprenticeship PD experiences are designed to model teacher pedagogies. Engaging teachers in analyzing the instructional design of specific PD experiences helps make design principles visible to teachers. After engaging in reading and tasks in Reading Apprenticeship PD, teachers analyze the instructional design of each learning experience, considering the level and degree of instructional support students would need to be successful. They examine and critique scaffolds and tools and offers suggestions to support any gaps or problems they identify.

Instructing and coaching. Teachers engage in learning in Reading Apprenticeship professional development, they are asked frequently throughout the sessions to make connections to their classrooms and students. This ongoing support for enacting new ideas and approaches in teachers’ own teaching and learning contexts can be brief—capturing thoughts for a few moments of writing—or involve extended instructional planning sessions.

Reflection on learning. Metacognition is central to the Reading Apprenticeship framework and aims to build learner agency and independence in the classroom. For teachers, routine reflections on what they are learning model the work for the classroom, but also engage teachers themselves in agentic and purposeful learning for their own knowledge and practice. Simultaneously, shared with professional development facilitators, reflections on learning offer formative assessment.

believe the study makes a timely contribution to theoretical understandings about the kind of PD required to integrate literacy and content area instruction.

5.1. Sample recruitment and selection

A number of important studies of literacy instruction have been conducted in the context of design-based research involving highly regarded teachers (e.g., Applebee, 1989; Langer, 1998; Newell, VanDerHeide, & Olsen, 2014). These studies are valuable not only for identifying promising practices, but also for surfacing stubborn obstacles to reform (Langer, 1998). In this tradition and with this broader research agenda in mind, teachers in both the PD and comparison groups were recruited and selected by research team members and educational leaders in schools and districts with whom they worked based on perceptions of good instructional practice and as potential long-term partners in our collaborative design work. Expecting these teachers to bring knowledge and expertise that could assist in the design of new approaches to teaching argumentation in their subject areas, we requested permission to observe their teaching. We also invited the teachers to participate in the larger design-based research project. Although participation in Reading Apprenticeship PD occurred prior to the current study and random assignment was not possible, teachers in the PD and control groups demonstrated comparable motivation and professionalism as evidenced by a willingness to have researchers observe and videotape their classrooms.

Teachers in both groups taught a wide variety of students in diverse settings as indicated by the percentage of students qualifying for free/reduced lunch in these classrooms, a proportion nearly identical to the national average (48% versus 50%). Characteristics of the PD and comparison teachers and lessons are shown in Table 2. Below we describe the recruitment and selection processes for the teachers in the PD and comparison groups.

**PD group.** The 16 teachers in the PD group were known to the research team because they had previously participated in 7-day Reading Apprenticeship professional development institutes and had been members of an ongoing professional development network of middle and high school teachers (the Continuing Network) from the greater San Francisco Bay Area. The years in which these teachers participated in the PD and the Continuing Network are shown in Table 3. Nearly two-thirds participated in

---

1 The magnitude of kappa is affected by the base rate of the target activity. When base rates for codes are uneven such that a few activities or interactions are more prevalent than others, as was the case with our coding scheme, large correction for chance agreement has a substantial impact and kappa values will be lower than when base rates are equal (Sim & Wright, 2005).

2 In Illinois, Type 03 (elementary) teaching certificates authorize the holder to teach grades K-9. In California, Elementary (Multiple Subject) Teaching Credentials authorize the holder to teach in any self-contained, core, or team teaching classroom.

3 One PD teacher held both an elementary/multiple subject and single subject credential.

4 One PD and one comparison teacher were observed teaching both ELA and History.

5 Four PD and two comparison teachers were observed teaching more than one grade.
Table 2
Teacher demographics.

<table>
<thead>
<tr>
<th>Demographic Variable</th>
<th>Categories</th>
<th>PD Group (N = 16)</th>
<th>Comparison Group (N = 18)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>14</td>
<td>88</td>
<td>12</td>
</tr>
<tr>
<td>Male</td>
<td>2</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Total years teaching</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 5 years</td>
<td>1</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>6–14 years</td>
<td>8</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>≥ 15 years</td>
<td>6</td>
<td>40</td>
<td>1</td>
</tr>
<tr>
<td>Number with Master’s degree</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Master’s degree</td>
<td>11</td>
<td>73</td>
<td>13</td>
</tr>
<tr>
<td>Teaching certification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Elementary</td>
<td>3</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Single subject</td>
<td>13</td>
<td>87</td>
<td>6</td>
</tr>
<tr>
<td>Subject(s) taught/observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ELA</td>
<td>8</td>
<td>50</td>
<td>7</td>
</tr>
<tr>
<td>History</td>
<td>7</td>
<td>44</td>
<td>5</td>
</tr>
<tr>
<td>Science</td>
<td>2</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Grade(s) taught/observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-8</td>
<td>6</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>9-12</td>
<td>11</td>
<td>69</td>
<td>13</td>
</tr>
<tr>
<td>Number lessons observed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>12</td>
<td>75</td>
<td>13</td>
</tr>
<tr>
<td>3-5</td>
<td>4</td>
<td>25</td>
<td>5</td>
</tr>
</tbody>
</table>

Note. Data are missing for total years teaching, degree, and certifications for one teacher in the PD group and three teachers in the comparison group. Consequently, percentages for these three variables are based on N’s of 15 for the PD group and 15 for the comparison group. All other percentages are based on N’s of 16 for the PD group and 18 for the comparison group.

Table 3
PD teachers’ participation in reading apprenticeship professional development and continuing network.

<table>
<thead>
<tr>
<th>Initial Reading Apprenticeship PD</th>
<th>(N = 16)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Before 2005</td>
<td>4</td>
</tr>
<tr>
<td>2005–2006</td>
<td>1</td>
</tr>
<tr>
<td>2006–2007</td>
<td>1</td>
</tr>
<tr>
<td>2007–2008</td>
<td>4</td>
</tr>
<tr>
<td>2008–2009</td>
<td>3</td>
</tr>
<tr>
<td>2009–2010</td>
<td>2</td>
</tr>
<tr>
<td>2010–2011</td>
<td>1</td>
</tr>
<tr>
<td>Years in Continuing Network</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>≥4</td>
<td>4</td>
</tr>
</tbody>
</table>

their initial Reading Apprenticeship PD at least two years before they were observed for this study; a quarter participated in their initial PD at least five years before the study. Most teachers in the PD group had been members of the Continuing Network for 1–2 years following their initial PD experience. Although teachers belonged to various PD cohorts, the design principles that shaped PD learning activities across cohorts are rooted in the theory and practices underlying the Reading Apprenticeship Framework as described above (Schoenbach et al., 2012). Therefore, regardless of when they participated, they shared a comparable PD experience.

In addition to observing their classrooms, we also invited PD teachers to participate in an ongoing teacher-researcher inquiry network in which the culture was deliberately one of co-learning among teachers and researchers. Thirteen of the 16 teachers accepted our invitation. During its first year, teachers met with researchers four times to read and discuss the grant proposal, Toulmin’s (1958) definition of argumentation, and texts and tasks from their own and others’ classrooms with an aim of understanding what text-based argumentation might be and what it might demand of students. Building common language and understanding over the first year enabled the work of the inquiry network to accelerate as the project began collaboratively designing, implementing, and refining argumentation interventions in the following years. Although our observations of teachers’ current practices overlapped with teachers’ participation in the inquiry network, the observations occurred prior to the design of any instructional intervention.

Comparison group. Teachers in the comparison group were nominated by research team members who had worked with Chicago area schools and teachers. Team members nominated teachers they believed to be engaging in instruction designed to foster disciplinary literacies and/or who were reported to have good instructional rapport resulting in high student engagement. We also solicited teacher nominations from district leadership in literacy, ELA, social sciences, and sciences. Because the sample of Reading Apprenticeship teachers included only two science teachers, we recruited two additional Bay Area science teachers who had not participated in Reading Apprenticeship PD but were recommended by colleagues as strong teachers of science. These science teachers were included in the comparison group.

As in the case of teachers in the PD group, comparison teachers were invited to participate in the larger research project. Sixteen of the 18 teachers accepted the invitation and served as disciplinary design team partners (4 teachers), members of a three-year Chicago teacher network (5 teachers, with 3 teachers participating all three years), or volunteered their classes to participate in piloting texts and tasks to inform the design of assessments of evidence-based argumentation (7 teachers).

5.2. Video data collection methods

We videotaped a total of 71 lessons. A single camera was placed at the rear of the room and focused on the teacher, although students were also visible much of the time. In addition to capturing opportunities to learn that were the target of our research, this protocol standardized the videotaping across classrooms (Jacobs, Hollingsworth, & Givvin, 2007). To explore promising practices, we revisited classrooms as warranted. Classrooms where teachers focused on building students’ capacities to carry out intellectually engaged reading and reasoning about text were visited more frequently than classrooms where students spent the majority of time in retrieval and reporting of textual information. We discuss how we addressed the differing number and length of videotaped lessons in the “Data Analysis” section, below. In total, we coded 3813 min of instruction, including 1944 min in PD classrooms and 1869 min in comparison classrooms.

6. Data analysis process

6.1. Coding scheme

Our coding captured opportunities presented by teachers for students to engage in activities, tasks, and interactions related to literacy learning. Although OTL has been measured in a variety of ways, our analysis focused on exposure, the amount of time teachers implemented or enacted various activities and tasks (Smith, 2000).

We examined three dimensions related to literacy OTL: 1. Content Delivery, 2. Task, and 3. Grouping. Definitions of the dimensions and individual codes are found in Table 4. Although we did not directly assess quality of instruction, our focus on cognitively
Content Delivery Dimension captures the mechanism by which students are taught lesson content.

1. **Teacher.** Content is delivered through teacher lecture, explanation, or PowerPoint, in which the teacher has done the work of understanding, organizing, and delivering material to students. This mode of content delivery may involve some student interaction around presented material, but the primary focus is on teacher-delivered content.

2. **Working with text.** Content is delivered through textual sources. This code is used when students work with text, individually or collaboratively, as a source of curriculum content. Text is defined broadly to include reading a wide range of materials, including graphic representations, from a wide range of sources, including computer screens, not just connected text and traditional print material. Working with text is distinguished from other Content Delivery codes in that it requires students to access the content provided in textual sources.

**Task Dimension** captures opportunities for students to engage in complex literacy and content learning tasks.

**Literacy tasks**

1. **Close reading.** Close reading tasks are characterized by approaching texts to understand and build meaning rather than only to find factual information. Although other definitions of close reading have been advanced, in our coding we defined close reading as active engagement in meaning-making with texts. These tasks ask students to engage in interpretation and argumentation processes to unearth and evaluate possible meanings at the local and global levels (Norris & Phillips, 2003), individually or collaboratively.

2. **Argumentation.** Based on Toulmin (1958), we define argumentation tasks as tasks that ask students to make a claim or assertion that is supported by evidence that connects to the claim in a principled way. This definition recognizes that argument is shaped by the discourses of particular communities of practice and consequently accommodates a wide range of tasks, from reason-giving interpretations of literature to modeling and explanation tasks in science. Tasks may or may not be explicitly identified as “argumentation” by the teacher.

3. **Cross-textual analysis.** These tasks ask students to synthesize, evaluate, or critique information from multiple texts.

4. **Disciplinary knowledge focus.** Disciplinary knowledge focus tasks ask students to engage with overarching frameworks, concepts, and/or themes of the discipline.

5. **Fact acquisition.** Fact acquisition tasks focus on testing understanding, recall, or rote learning with little or no opportunity for sense-making. Although they may involve brief instructional exchanges between teacher and students, the overwhelming focus is on learning facts/information and right answers.

**Grouping Dimension** captures opportunities for different kinds of learning interactions and supports.

1. **Individual.** Students work independently.

2. **Pairs.** Students work in pairs.

3. **Small group.** Students are divided into small groups that they generally run themselves.

4. **Whole class.** Teacher interacts with the whole class at once.

**6.2. Coding protocol**

Because we operationalized OTL as exposure, we coded the duration of each coded activity. In addition to yielding absolute durations, this approach enabled us to calculate percentages for comparing observations of varying numbers and lengths (Garnier, Lemmens, Druker, & Roth, 2011). Because classrooms are complex environments where multiple activities often occur simultaneously, most footage was multiply coded. Consequently, percent time within a dimension does not sum to 100.

**Coder training.** Video data were coded by four researchers, all of whom had classroom teaching experience. Coding occurred prior to the decision to compare PD and comparison teachers; therefore, bias based on knowledge of group membership was unlikely. Furthermore, two coders were members of the Bay Area research team, site of PD group observations, and two were members of the Chicago research team, site of most comparison group observations. All four coders were assigned observations from both sites.

Researchers coded directly from video footage using NVivo9 qualitative analysis software (QSR International), without transcripts of the lessons. Training and support for coding was provided by the authors. Trainings employed a frame of reference model based on practice, discussions between coders, and feedback from trainers (Melchers, Lienhardt, Von Aarburg, & Kleinmann, 2011). Once we established acceptable inter-rater reliability among the four coders, each video was assigned and coded by a single coder. To insure and maintain good inter-rater reliability, two or three additional videotapes were coded and discussed by pairs of coders throughout the coding phase. Ten percent of videos were multiply coded.

**Inter-rater reliability.** After calculating kappa coefficients for each jointly coded video, we calculated a generalized version of the kappa statistic for each code by averaging kappas across the seven videos coded by multiple coders. Average kappas indicated almost perfect (0.81–0.99) or substantial (0.61–0.80) agreement for 7 of the 11 codes (Viera & Garrett, 2005). Two codes showed moderate agreement (0.41–0.60) and two showed fair agreement (>0.30 to 0.40). Percentage agreement for 8 of the 11 codes was between 85% and 97%, and was 75% or above for the other three codes.

**6.3. Data analysis**

Because both the number and length of videotaped lessons differed for different teachers, following the Third International Mathematics and Science Study (TIMSS) videotape classroom studies (Hiebert et al., 2003), we transformed duration codes into mean percent of class time variables. As a first step, we computed percent of class time variables for individual teachers. In the next step, we used teacher percent of class time variables to calculate mean percent of class time separately for PD and comparison groups. We also calculated interquartile ranges to look at the distribution of time allocated to various activities by PD and comparison teachers.

In addition to exploring patterns of practice based on these numerical data, we conducted independent samples t-tests to compare mean percent class time allocated by PD and comparison teachers. Based on Reading Apprenticeship PD emphases, we predicted that PD teachers would provide a greater literacy focus and provide greater opportunity for students to do the intellectual work of learning. Because Reading Apprenticeship PD focuses on building teachers’ capacity to tightly integrate literacy and content learning, we did not expect increased literacy learning opportunities to occur at the expense of content learning. Therefore, we predicted no difference between PD and comparison teachers in opportunities for students to learn content.
7. Results

Mean percent of class time for content delivery, task, and grouping codes is shown in Table 5. Below we explore patterns of practice revealed by these data.

7.1. The centrality of text

Our first research question focused on opportunities provided by teachers for students to work with text, in contrast to learning content through teacher lecture or explanation. Nearly every teacher lectured and explained content and allocated time for working with text. Nonetheless, as shown in Table 5, we found group differences in the mean percent of class time allocated to the two content delivery modes. Teachers in the PD group allocated considerably less time to delivering content on average than comparison teachers and more time to working with text.

7.2. A focus on complex literacies

Our second research question focused on opportunities provided by teachers for students to engage in complex literacy practices. As shown in Table 5, PD teachers allocated more time both to close reading, a central focus of Reading Apprenticeship PD, and to argumentation and cross-textual analysis, complex literacy practices not explicitly addressed by Reading Apprenticeship PD but emphasized by literacy standards emerging at the time the data were collected.

7.3. Integration of literacy and content learning

Research Question 3 queried the relationship between literacy and content learning. Our analysis of teachers’ integration of literacy and content learning was based on overlap from video footage that was co-coded to literacy and content learning tasks. Concern has been raised that attention to literacy may diminish the quality of content learning (Misulis, 2009). To consider this possibility, we looked at the extent to which literacy tasks focused on building disciplinary knowledge. As shown in Fig. 1, literacy tasks were consistently associated with a higher than average disciplinary knowledge focus in PD classrooms. Although PD teachers allocated 46% of class time overall to tasks with a disciplinary knowledge focus, close reading in PD classrooms had a disciplinary knowledge focus 52% of the time, argumentation had a disciplinary knowledge focus 59% of the time, and cross-textual analysis had a disciplinary knowledge focus 84% of the time. In comparison classrooms, argumentation and cross-textual analysis, though rare (see Table 5), were likewise associated with a higher than average disciplinary knowledge focus. However, close reading was co-coded with disciplinary knowledge 37% of the time—considerably below the 58% of time comparison teachers allocated overall to disciplinary knowledge tasks. This suggests that whereas reading and reasoning served disciplinary knowledge building in PD classrooms, in comparison classrooms, reading resulted in a reduced focus on disciplinary learning.

Table 5
Total minutes and mean percent class time allocated to content delivery, task, and grouping opportunity to learn.

<table>
<thead>
<tr>
<th></th>
<th>Total Teachers (N = 34)</th>
<th>3813 min</th>
<th>PD (N = 16)</th>
<th>1944 min</th>
<th>Comparison (N = 18)</th>
<th>1869 min</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content Delivery</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher</td>
<td>181</td>
<td>12</td>
<td>553</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working with text</td>
<td>1419</td>
<td>65</td>
<td>865</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Task</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argumentation</td>
<td>346</td>
<td>13</td>
<td>173</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Close reading</td>
<td>850</td>
<td>35</td>
<td>200</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cross-textual analysis</td>
<td>205</td>
<td>9</td>
<td>78</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disciplinary knowledge focus</td>
<td>1016</td>
<td>46</td>
<td>999</td>
<td>58</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fact acquisition</td>
<td>146</td>
<td>14</td>
<td>462</td>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Grouping</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual</td>
<td>363</td>
<td>17</td>
<td>257</td>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pairs</td>
<td>339</td>
<td>17</td>
<td>118</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small group</td>
<td>307</td>
<td>17</td>
<td>316</td>
<td>19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole class</td>
<td>886</td>
<td>44</td>
<td>1129</td>
<td>56</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Content Delivery and Task categories do not sum to 100% of total lesson time for several reasons: students sometimes learned lesson content in ways other than teacher lecture/explanation or working with text (e.g., through media, labs, or simulations); students were often asked to perform multiple tasks during a single segment so most footage was multiply coded; and time allocated to housekeeping was not coded to any content delivery or task.
We also looked at tasks students were asked to perform when working with text. As shown in Fig. 2, although working with text in PD classrooms was associated with high levels of content learning, it also frequently engaged students in close reading and, to a lesser extent, argumentation and cross-textual analysis. In contrast, working with text in comparison classrooms overwhelmingly focused on content learning and only rarely involved close reading or reasoning tasks. Thus, even working with text was no guarantee of a literacy focus in comparison classrooms.

7.4. Shifting the intellectual work of learning to students

Our fourth research question focused on grouping as an indicator of opportunity for students to participate in various kinds of learning interactions and supports. As shown in Table 5, PD teachers allocated class time more evenly across grouping structures than comparison teachers. In particular, PD teachers allocated less time to whole class instruction and more time to working in pairs than comparison teachers.

Because we were particularly interested in interactions and support for literacy learning, we examined relationships between grouping and literacy OTL. In general, differences between PD and comparison teachers were most pronounced for peer-directed and whole class instruction. As shown in Table 6, peer-directed learning in both PD and comparison classrooms generally involved working with text. However, although nearly half of the time allocated to pair and small group work in PD classrooms involved close reading, pairs and small groups in comparison classrooms rarely involved close reading. There were also differences in how PD and comparison teachers used whole class instruction. PD teachers spent less whole class time lecturing and explaining and more whole class time working with text than comparison teachers. Similarly, PD teachers devoted considerable whole class instruction to literacy tasks, particularly close reading. In contrast, comparison teachers used whole class instruction almost exclusively for content acquisition.

7.5. Variability within groups

Table 7 reports the distribution of these practices using interquartile range. As noted above, PD and comparison teachers offered substantially different OTL to their students, with PD teachers allocating less time to lecturing and explaining and more time to complex literacy practices and partner interactions than comparison teachers. Interquartile ranges for content delivery, task, and grouping OTL support and extend these findings. For example, the narrow interquartile range for teacher delivery of content among members of the PD group indicates that not only did PD teachers allocate less time on average to lecturing and explaining, but this was true of the bulk of PD teachers. In contrast, a wider interquartile range for comparison teachers indicates that there was considerable variation among comparison teachers in this practice. A similar pattern is seen in time allocated to fact acquisition.

Comparison teachers exhibited a narrow interquartile range for argumentation, close reading, and cross-textual analysis compared with PD teachers. An examination of quartile values indicates that this was due to a dearth of these complex literacy practices in comparison classrooms. Similarly, interquartile range indicates that there was little pair work in comparison classrooms.

7.6. Statistical contrasts between PD and comparison teachers

Table 8 reports results of independent samples t-tests. As predicted, PD teachers had a greater literacy focus, allocating significantly less time to delivering content by lecturing and explaining ($t = 2.59, p = .015$), and more time to working with text ($t = 2.37, p = .024$) than comparison teachers. PD teachers also devoted significantly more time to close reading than comparison teachers ($t = 3.16, p = .005$). Although patterns of practice indicated that PD teachers allocated considerably more time on average to argumentation and cross-textual analysis (see Table 5), mean differences between PD and comparison teachers were not significant for these literacy practices that were not explicitly addressed by Reading Apprenticeship PD at the time.

As predicted, students in PD classrooms had greater opportunity to participate in peer-directed learning. Specifically, PD teachers allocated more time for students to interact in pairs than comparison teachers ($t = 2.13, p = .041$). There was no significant difference between PD and comparison teachers in time allocated to small groups.

Finally, as predicted, there were no differences between PD and comparison teachers on time allocated to tasks focused on content learning, either in the form of disciplinary knowledge tasks or fact acquisition tasks. This supports our hypothesis that the increased literacy focus in PD classrooms did not result in diminished opportunity to learn course content.

8. Discussion

Our study found differences between PD and comparison teachers in practices that reflected core elements of the Reading Apprenticeship instructional framework and PD model: (1) the centrality of texts to the authentic intellectual work of learning, (2) a focus on complex literacies, (3) deeply integrating these literacy learning opportunities with content learning, and (4) transferring ownership of making meaning to students through opportunities for social collaboration and scaffolding from both teachers and peers. These elements are likewise emphasized by current literacy reforms, including the Common Core, C3 Framework for Social Studies State Standards, and Next Generation Science Standards. Below we briefly discuss the impact of Reading Apprenticeship PD on these literacy practices before turning to implications for PD with the potential to build teachers’ capacity for integrating literacy and content learning.
often preferred to other measures of variability, such as standard deviation, because it is more robust to outliers and non-normal data. The interquartile range (IQR) is calculated as the difference between the third quartile (Q3) and the first quartile (Q1) of a data set. It represents the range within which the middle 50% of observations fall.

Note: Interquartile range describes the middle 50% of observations in a data set. If the interquartile range is large, it means that the middle 50% of observations are widely spaced. If the interquartile range is small, it means that the middle 50% of observations are narrowly spaced. Because it focuses on the middle 50%, the interquartile range is less sensitive to outliers and non-normal data and is, therefore, often preferred to other measures of variability, such as standard deviation (Manikandan, 2011).

### Table 7
Interquartile range for percent time allocated to content delivery, task, and grouping opportunity to learn.

<table>
<thead>
<tr>
<th>Content Delivery</th>
<th>PD, n = 16</th>
<th>Comparison, n = 18</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Q1-Q3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teacher</td>
<td>10.8% (15.0–43.3%)</td>
<td>27.0% (38.7–117.7%)</td>
</tr>
<tr>
<td>Working with text</td>
<td>24.3% (79.8–55.5%)</td>
<td>27.9% (99.8–51.9%)</td>
</tr>
<tr>
<td>Task</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argumentation</td>
<td>22.1% (22.1–0.0%)</td>
<td>67.0% (0.0–0.0%)</td>
</tr>
<tr>
<td>Close reading</td>
<td>54.4% (54.4–0.0%)</td>
<td>65.0% (6.5–0.0%)</td>
</tr>
<tr>
<td>Cross-textual analysis</td>
<td>20.7% (20.7–0.0%)</td>
<td>20.8% (0.0–0.0%)</td>
</tr>
<tr>
<td>Disciplinary knowledge focus</td>
<td>31.3% (32.3–63.1%)</td>
<td>31.3% (77.4–45.7%)</td>
</tr>
<tr>
<td>Fact acquisition</td>
<td>10.4% (10.4–0.0%)</td>
<td>22.7% (25.2–2.6%)</td>
</tr>
<tr>
<td>Grouping</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouping: Pairs</td>
<td>23.4% (27.1–3.7%)</td>
<td>60.0% (6.6–0.0%)</td>
</tr>
<tr>
<td>Grouping: Small groups</td>
<td>36.3% (36.3–0.0%)</td>
<td>35.8% (35.8–0.0%)</td>
</tr>
</tbody>
</table>

Note. Interquartile range describes the middle 50% of observations in a data set. If the interquartile range is large, it means that the middle 50% of observations are widely spaced. If the interquartile range is small, it means that the middle 50% of observations are narrowly spaced. Because it focuses on the middle 50%, the interquartile range is less sensitive to outliers and non-normal data and is, therefore, often preferred to other measures of variability, such as standard deviation (Manikandan, 2011).

### 8.1. Impact on reading apprenticeship PD literacy practices

**The centrality of texts.** The centrality of texts observed in PD classrooms is rare in U.S. secondary schools (Greenleaf & Valencia, 2016). Research indicates that teachers in all subject areas and academic tracks use lecture as the predominant mode of instruction (ACT, Inc., 2006; Fisher, 2009; Ness, 2008). Although teachers report assigning significant reading (ACT, Inc., 2013a; Opfer et al., 2016), students themselves may do little of it. Instead, teachers often resort to showing or telling students about content as an efficient alternative to actively engaging students in making sense of challenging academic texts (Santelises & Dobrowski, 2015; Santelises & Dobrowski, 2016). The fact that PD teachers allocated 12% of class time to lecturing and explaining and 65% to working with text thus represents a dramatic shift from traditional content area instruction dominated by teacher lecture toward text-based instruction mandated by current literacy reform initiatives.

**A focus on complex literacies.** Our study set a high bar for gauging the impact of Reading Apprenticeship PD on opportunities for students to engage in complex literacies. In addition to close reading, a major focus of the PD, we looked at time teachers

### Table 8
Independent samples test for content delivery, task, and grouping opportunity to learn.

<table>
<thead>
<tr>
<th>Content Delivery</th>
<th>PD, n = 16</th>
<th>Comparison, n = 18</th>
<th>t-test for Equality of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M</strong></td>
<td>16.4%</td>
<td>26.9%</td>
<td>20.3%</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>15.9%</td>
<td>23.5%</td>
<td>22.1%</td>
</tr>
<tr>
<td>Task</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Argumentation</td>
<td>17.2%</td>
<td>17.2%</td>
<td>147.7%</td>
</tr>
<tr>
<td>Close reading</td>
<td>31.9%</td>
<td>17.2%</td>
<td>21.2%</td>
</tr>
<tr>
<td>Cross-textual analysis</td>
<td>13.5%</td>
<td>58.3%</td>
<td>26.1%</td>
</tr>
<tr>
<td>Disciplinary knowledge focus</td>
<td>22.5%</td>
<td>20.1%</td>
<td>21.2%</td>
</tr>
<tr>
<td>Fact acquisition</td>
<td>16.8%</td>
<td>6.3%</td>
<td>12.4%</td>
</tr>
<tr>
<td>Grouping</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grouping: Pairs</td>
<td>17.0%</td>
<td>19.0%</td>
<td>20.7%</td>
</tr>
<tr>
<td>Grouping: Small groups</td>
<td>17.0%</td>
<td>18.1%</td>
<td>21.1%</td>
</tr>
</tbody>
</table>

Note. *p < .05, **p < .01. Bold values represent independent samples test demonstrated statistically significant differences between the two groups. * Results support prediction of no difference between PD and comparison teachers.
devoted to argumentation and cross-textual analysis, literacy practices on the reform horizon in 2010–2011 when these data were collected (NGA & CCSSO, 2010). Our findings suggest that participation in Reading Apprenticeship PD not only generated a sustained emphasis on close reading—PD teachers devoted 35% of class time to close reading—but also primed teachers for other complex literacy practices emphasized by new literacy standards. It may be that close reading with its focus on evaluating different possible interpretations of text provided a bridge to reform efforts calling for students to use evidence from multiple sources to interpret literature, history, and science (NCSS, 2013; NGSS Lead States, 2013; NGA & CCSSO, 2010). An analysis of co-occurrence among literacy tasks indicated that these complex literacies often occurred together in PD classrooms, but not in comparison classrooms.

**Capacity to integrate literacy and content learning.** Our results challenge the notion that literacy and content learning are in competition. Although PD teachers devoted substantially more time to close reading and other complex literacies, this did not occur at the expense of disciplinary knowledge building. There was no difference between PD and comparison classrooms in time allocated to tasks with a content learning focus. Furthermore, literacy tasks in PD classrooms frequently co-occurred with more cognitively demanding disciplinary knowledge tasks.

At the same time, our findings underscore the need for PD that builds teachers’ capacity to integrate literacy and content learning. In contrast to the balance between literacy and content learning observed in PD classrooms, instruction in comparison classrooms was skewed toward content acquisition. Even when students were ostensibly working with text, they rarely engaged in the kind of close academic reading required for disciplinary knowledge building (Norris & Phillips, 2003). Furthermore, despite a strong emphasis on acquiring content, close reading in comparison classrooms was associated with a decreased disciplinary knowledge focus. Our findings thus corroborate research indicating that, absent intervention, teachers have considerable difficulty integrating literacy and content learning (Greenleaf & Valencia, 2016; Misulis, 2009; Reed, 2009).

**Shifting the intellectual work of learning to students.** Patterns of practice suggest that PD teachers provided both more instructional support and greater opportunity for students to engage in self-directed learning than comparison teachers, particularly pair work, which provides greater opportunity for active participation than group work (Dobao and Blum, 2013). This is noteworthy. Research suggests that shifting interpretive authority to students is particularly challenging for teachers (Applebee, Langer, Nystrand, & Gamoran, 2003; Santelieses & Dobrowski, 2015; Smith, Hardman, Wall, & Mroz, 2004). A recent study of implementation of the Common Core found that teachers overwhelmingly retained interpretive control, providing only brief moments for students to talk together about texts (Santelieses & Dobrowski, 2015).

Although our coding did not permit us to look at specific interactions associated with grouping patterns, PD teachers provided ample opportunities for student-student talk—over a third of class time was spent in peer-led groups, compared to the brief moments of student talk described by Santelieses and Dobrowski (2015). Furthermore, an examination of individual lessons revealed that whole class instruction in PD lessons frequently alternated with individual and peer-directed groups. Importantly, this recursive cycle is embedded in every Reading Apprenticeship PD inquiry, as illustrated by the Superman and Me inquiry above. In contrast, peer-directed groups in comparison classrooms were of longer duration and less frequently punctuated by whole class instruction. A wealth of research demonstrates that productive instructionally-focused conversation among students is difficult to achieve and maintain and that simply putting students together is not enough (Murphy, Wilkinson, & Soter, 2011). This suggests that shifting the intellectual work of learning to students may be supported by the type of intermittent whole class teacher-facilitation modeled in Reading Apprenticeship PD and subsequently enacted in many PD classrooms we observed.

### 8.2. Implications for professional development design

A study of factors influencing the implementation and sustainability of federally funded educational innovations from the 1950s and 60s found that initiatives that actively engaged teachers in integrating reform strategies with current classroom practices were more likely to endure than projects that were overly planned or prescribed (Berman & McLaughlin, 1978). Furthermore, given time and opportunity to make their own sense of new ideas, ambitious reform efforts requiring significant change in teacher practice were more likely to endure than innovations of more limited scope. Although these understandings originated 40 years ago and have garnered additional support in ensuing years, an exchange in Educational Researcher over the kind of PD needed to support literacy reform demonstrates that the issue is far from resolved (Anderson & Herr, 2011; Baumann & Barry, 2011). Referencing scale-up pressures related to implementing the Common Core, Baumann and Barry (2011) challenge the efficacy of efforts to translate the extensive research base on teacher professional learning into PD methods, concluding that “it is likely because these concepts are complex and take time to implement, that a broad scale-up of professional development of this kind to teachers nationally would present challenges” (p. 176). Although scale-up pressures cited by Baumann and Barry are real (Kaufman et al., 2016), research provides ample evidence that shortchanging an investment in teachers’ deep learning is a major contributor to the failure of reform efforts (Elmore, 1996). Without PD that actively supports re-enacting as well as re-understanding, experience suggests that instruction “changes very little, except in a small proportion of schools and classrooms where the changes do not persist for very long” (Elmore, 1996, p. 7).

**Role of inquiry.** Researchers have long promoted an inquiry stance toward teacher education and PD, with teachers learning in and from practice by documenting and reflecting on their work and the responses of their students (Cochran-Smith & Lytle, 1996). Yet, curiously, inquiry is not among the five widely embraced core features of effective PD (Desimone, 2009). This omission may be an artifact of the frequent concurrence of inquiry pedagogies with PD models aimed at developing teachers’ understanding of what they teach and how students learn that content—or pedagogical content knowledge (PCK) (Shulman, 1986)—considered by some researchers to be the most important element of effective PD (e.g., Desimone, 2009; Guskey & Yoon, 2009; Kennedy, 1998; Reed, 2009; Yoon, Duncan, Lee, Scarluss, & Shapley, 2007). Because PCK sufficient to improve teaching and learning requires deep understanding, PCK-focused PD is often designed around inquiry activities that draw on authentic aspects of classroom practice through exploring records of practice, such as student work samples or video recordings of classroom instruction, and opportunities to enact and reflect on target pedagogies, both individually and with others. As McNeill and Knight (2013) note, these inquiry pedagogies are more likely to produce deep, generative knowledge than PD employing more traditional pedagogies. Although McNeill and Knight specifically address PD aimed at improving PCK, Kennedy (2016) found that this was true whether PD focused on PCK or other problems of practice. Inquiry is thus a frequent behind-the-scenes player in effective PD, and inquiry pedagogy likely
The Reading Apprenticeship PD experienced by teachers in the current study incorporated all five core features of Desimone’s (2009) PD framework. Thus, factors other than inquiry likely contributed to differences we observed between PD and comparison teachers in the current study, including teachers’ participation in the Continuing Network. However, a study by Garet et al. (2008) suggests that absent inquiry pedagogy, Desimone’s (2009) framework may not be sufficient to produce the kind of lasting impact on literacy instruction that we observed in PD classrooms in our study. Teachers in the Garet et al. (2008) study received eight days of PD designed to improve PCK around literacy. Rather than the inquiry pedagogy characteristic of most PCK-focused PD, the core of the training was an ongoing PowerPoint presentation. To align the PD with Desimone’s five-part framework, segments of the PowerPoint were interspersed with open-floor discussions and opportunities for active participation. Despite these efforts, the PD produced only fleeting changes in teacher knowledge and classroom practice and no impact on student achievement.

Furthermore, findings from Garet et al. (2008) suggest that absent inquiry pedagogy, continuing support for teachers may not improve or sustain the impact of PD. Half the teachers in Garet et al. (2008) received intensive coaching following the PD institute—two hours per week per teacher for a total of 60 h over the academic year. Yet like the PD-alone condition, PD-plus-coaching failed to produce changes in classroom practice or student outcomes. It is worth noting that the Continuing Network in the current study involved a fraction of the time devoted to one-on-one coaching in Garet et al. (2008)—meeting just a few times a year to continue inquiry into literacy teaching and learning. Thus, although the Continuing Network may have contributed to the sustainability of Reading Apprenticeship practices, the inquiry nature of continuing support, rather than continuing support per se, was likely a critical factor.

Inquiry PD for complex literacies. Based on its enduring effect on teachers’ practice, we believe the Reading Apprenticeship PD experienced by teachers in this study provides a model of the type of PD needed to build teachers’ capacity for improving literacy instruction across the content areas. Importantly, although inquiry is central to this model, in recognition of the many demands on teachers’ time and attention, inquiries are strategically focused to work as efficiently as possible to develop teacher capacities (Greenleaf & Schoenbach, 2004). Because it has been developed in situ by engaging diverse teacher and school communities over time, in changing policy environments, variations in teachers’ beliefs, knowledge, and experiences and teaching contexts have significantly informed the PD design. Although initially developed and continually refined through close, collaborative design research partnerships, the PD has been provided to teachers on a large scale, with the support of federal research and dissemination grants (Fancsali et al., 2015; Greenleaf et al., 2011; Somers et al., 2010).

To build teachers’ enduring capacity to teach complex literacies alongside content, as mandated by current literacy reform efforts, our experience suggests that PD must engage teachers in inquiry-based learning experiences that:

- target specific teacher learning outcomes vital to the transformation of classroom learning environments;
- are enacted as models of instructional practices and pedagogies it is meant to foster;
- open spaces for teachers to respectfully voice and interactively interrogate multiple perspectives on texts, literacy, student performances, and instructional strategies;
- recognize and draw on teachers’ passions and expertise, while building new knowledge, skill, and motivations for change;
- assume and build teacher capacity to flexibly adapt ideas and practices to their specific circumstances;
- although goal-driven, maintain an active, inquiry orientation to support deeply experiential learning; and
- invite ongoing reflection on teaching and learning.

As indicated by findings from the current study, through such generative, inquiry-based PD, teachers gain enduring professional insights to guide their instruction over time and the capacity to flexibly and adaptably respond to the dynamic and ongoing flow of teaching (Gutiérrez & Penuel, 2014; Penuel, Fishman, Cheng, & Sabelli, 2011; Voogt et al., 2015). This is the kind of expertise we suggest will be required to address the new literacy standards, if we are to raise the learning and achievement and life chances of all students.

8.3. Limitations of the study

Teachers who participated in this study were selected because we believed their literacy implementation held some promise to inform our understanding of what is required to support complex literacies in ELA, history, or science. Therefore, literacy learning opportunities we observed may not represent opportunities on offer in most content area classrooms. Indeed, they may constitute a best-case scenario for students’ opportunity to learn. Nonetheless, we believe that what we learned has implications for the kind of PD needed to shift literacy instruction in more typical classrooms. Scale-up studies of Reading Apprenticeship PD consistently document positive impacts on both teacher and student outcomes among diverse groups of teachers. In addition, a robust body of literature indicates that traditional conceptions of knowledge, teaching, and learning may cause even highly motivated teachers to unconsciously assimilate curricular innovations into existing beliefs and practices (e.g., Gresalfi, Barnes, & Cross, 2012; Langer, 1998; Remillard, 2005; Simon, Erduran, & Osborne, 2006; Spillane et al., 2002). Thus, we believe it is likely that shifting the literacy practices of all teachers—both highly regarded and more typical teachers—will require substantial investments in teacher learning, with ample opportunities for both re-understanding and re-negotiating literacy beliefs and practices.

Our study design offered a tool to investigate whether participation in inquiry-based PD was related to differences between these groups of teachers, yet it is difficult to be certain whether observed differences were due to teachers’ prior experiences of Reading Apprenticeship PD or to confounding variables, such as geographic differences. In addition, as indicated by Table 2, demographic differences between PD and comparison teachers, including certification and teaching experience, may have contributed to the differences in teacher practices that we observed. Some reassurance is provided by the fact that literacy practices of Bay Area science teachers who were members of the comparison group appeared to more closely resemble the Chicago teachers than the Reading Apprenticeship PD cohort. In addition, our study results are bolstered by previous randomized controlled studies showing that participation in Reading Apprenticeship PD can have large and significant impacts on the classroom practices of subject area teachers and, therefore, their students’ learning and achievement in the year following the PD year (Greenleaf et al., 2011). The current study extends these findings by suggesting that these impacts endure—possibly for years after participation in...
PD. Furthermore, our finding that PD teachers provided greater opportunity for students to engage in complex literacy practices that were not the explicit focus of the PD suggests more distal impacts on literacy practices than those measured by previous studies. Finally, we know that opportunities teachers create constrain what students learn. Nonetheless, in focusing on opportunities presented by the teacher, our research presents only half the picture. Although dimensions of OTL emphasized by Reading Apprenticeship PD were informed by extant research about instruction that influences student learning outcomes, the current study cannot tell us how different opportunities provided by PD and comparison teachers influenced student learning. Happily, these observations of PD and comparison teachers informed a subsequent efficacy study based on the Reading Apprenticeship inquiry-based model of teacher PD that investigated student outcomes related to literacy practices that were the focus of the current study (Goldman et al., 2016; Greenleaf et al., 2016).

9. Conclusions

This study was motivated by the need for continuing research on the kinds of PD experiences that build teachers’ capacity to foster students’ in-depth understanding and engagement with literary and subject area learning. Although we acknowledge that “going to scale with effective PD for the entire teacher workforce of 3.7 million will require more research” (Wilson, 2013, p. 312), it is equally important to recognize that a large body of knowledge points to the kind of PD required to support the intellectually ambitious literacy goals of educational reform efforts such as the Common Core. Even researchers who argue that teacher learning is affected by a multitude of individual and contextual factors acknowledge that “there are generalizations that we should be able to make about the way professional learning activities relate to teacher learning that are true across different teachers and different school contexts” (Opper & Pedder, 2011, p. 394). Thus, although additional research and research designs are needed to help determine how to best support teachers (Hill, Beisiegel, & Jacob, 2013), the will to provide the kind of inquiry-based PD that is already known to be capable of transforming teachers’ beliefs, knowledge, and practices may be the major barrier to meaningful and enduring literacy reform (Darling-Hammond & McLaughlin, 1995). We believe that making headway at the scale necessary to provide all students with the kind of literacy instruction envisioned by recent reforms will necessitate providing all teachers with the kind of PD experience that has an enduring power to transform classroom instruction in secondary subject area classrooms. As studies of subject area teaching and school reform in middle and high school classrooms have long demonstrated, to proceed without such transformative learning opportunities for teachers will be a guarantee of failure (Cobb, Jackson, Smith, & Henrick, 2013; Elmore, 1996; Gutiérrez & Penneu, 2014; Knapp, 1997; Penneu et al., 2011; Rose, 2015).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.tate.2018.01.006.

References


ACT, Inc. (2013a). ACT national curriculum survey 2012: English language arts, Iowa City, IA.


