American Educational Research Journal
June 2011, Vol. 48, No. 3, pp. 647–717
DOI: 10.3102/0002831210384839
© 2011 AERA. http://aerj.aera.net

Integrating Literacy and Science in Biology: Teaching and Learning Impacts of Reading Apprenticeship Professional Development

Cynthia L. Greenleaf
Cindy Litman
Thomas L. Hanson
Rachel Rosen
WestEd
Christy K. Boscardin
Joan Herman
University of California-Los Angeles
Steven A. Schneider
WestEd
with Sarah Madden
WestEd
and Barbara Jones
University of California-Los Angeles

This study examined the effects of professional development integrating academic literacy and biology instruction on science teachers' instructional practices and students' achievement in science and literacy. The intervention consisted of 10 days of professional development in Reading Apprenticeship, an instructional framework integrating metacognitive inquiry routines into subject-area instruction to make explicit the tacit reasoning processes, problem-solving strategies, and textual features that shape literacy practices in academic disciplines. The study utilized a group-randomized, experimental design and multiple measures of teacher implementation and student learning and targeted groups historically unrepresented in the sciences. Hierarchical linear modeling procedures were used to estimate program impacts. Intervention teachers demonstrated increased support for science literacy learning and use of metacognitive inquiry routines, reading comprehension instruction, and collaborative learning structures compared to controls. Students in treatment classrooms performed better than controls on state standardized assessments in English language arts, reading comprehension, and biology.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

KEYWORDS: disciplinary literacy, science literacy, academic literacy, reading comprehension, professional development, STEM

Our democracy and future economic well-being depend on a literate populace, capable of fully participating in the demands of the 21st century (Carnegie Council on Advancing Adolescent Literacy [CCAAL], 2010; Rutherford & Ahlgren, 1990). Yet National Assessment of Educational Progress (NAEP) results indicate that most American youth lack the skills to successfully engage in the higher-level literacy, reasoning, and inquiry needed for an information-generating and information-transforming economy (NAEP, 2006, 2007, 2009; Snipes & Horwitz, 2008). Further, a persistent achievement gap persists between mainstream populations and those

---

CYNTHIA L. GREENLEAF is codirector of the Strategic Literacy Initiative at WestEd, 300 Lakeside Drive, 15th Floor, Oakland, CA 94612; e-mail: *cgreenl@wested.org*. Her research interests include professional development design, academic literacy instruction for underprepared youth, and discipline-specific literacy practices.

CINDY LITMAN is a senior research associate in the Strategic Literacy Initiative at WestEd; e-mail: *clitman@wested.org*. Her research interests include literacy development and relationships among social, emotional, and intellectual development.

THOMAS L. HANSON is a senior research associate at WestEd; e-mail: *thanson@wested.org*. His research interests include research methodology, academic achievement, and youth development.

RACHEL ROSEN is a doctoral candidate in the Leadership, Policy, and Politics program at Teachers College, Columbia University; e-mail: *rbr2111@columbia.edu*. Her current research interests are teacher quality and teacher labor markets.

CHRISTY K. BOSCARDIN is an assistant professor at the University of California, San Francisco, Department of Medicine and a member of the Office of Medical Education; e-mail: *christy.boscardin@ucsf.edu*. Her current research interests include medical education, assessment and measurement, and research methods and applied statistics.

JOAN HERMAN is the director of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles; e-mail: *herman@cse.ucla.edu*. Her research interests include the design and use of assessment to support improvement, the effects of accountability on schools and teaching, and evaluation of reform.

STEVEN A. SCHNEIDER is the senior program director of Science, Technology, Engineering, and Mathematics (STEM) at WestEd; e-mail: *sschnei@wested.org*. His research focuses on the efficacy and impact of STEM interventions.

SARAH MADDEN is a research associate for Science, Technology, Engineering, and Mathematics (STEM) at WestEd; e-mail: *smadden@wested.org*. Her interests are in environmental policy research.

BARBARA JONES is a research associate at the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), at the University of California, Los Angeles; e-mail: *bjones@cse.ucla.edu*. Her research interest is the use of assessment to understand and improve student learning.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

outside of the mainstream (Donahue, Voelkl, Campbell, & Mazzeo, 1999; Gee, 1999; Jencks & Phillips, 1998; J. Lee, Grigg, & Donahue, 2007; Mullis et al., 1994). Achievement gaps among different populations of students in reading are echoed by similar gaps in science learning and achievement (Grigg, Lauko, & Brockway, 2006; National Center for Education Statistics, 2007). The need for instructional support for adolescents' literacy development throughout schooling, though long neglected, has recently gained acknowledgement and support through state and national funding and policy mandates (CCAAL, 2010). As a result, in the current policy environment, schools and districts are under increased pressure to place low-performing students into reading intervention programs with the well-intentioned goal of increasing their literacy proficiency, yet some researchers suggest that the skill-based instruction they receive may perpetuate low literacy achievement rather than accelerate literacy growth (Allington & McGill-Franzen, 1989; Greenleaf & Hinchman, 2009; Haycock, 2001; Hiebert, 1991; Hull & Rose, 1989; C. D. Lee & Spratley, 2010). Moreover, intervention programs designed to address reading problems increasingly result in lost opportunities to learn in other academic subjects, particularly, science (McMurrer, 2007; Rentner et al., 2006). Withdrawing adolescents from instruction in science to remediate reading difficulties threatens to exacerbate historic inequities in achievement for populations of students traditionally underrepresented in the sciences (Barton, 2003). There is therefore increasing urgency to investigate how the integration of reading instruction into science learning at the high school level might advance the reading and science achievement of underachieving youth.

In this article, we advance the idea that we must think strategically about the integration of literacy development across subject matter domains if we expect to develop students' multiple capacities, particularly for students from groups who have been historically underrepresented in the sciences. Science classrooms can contribute significant opportunities for students to acquire greater literacy proficiency, just as greater literacy proficiency is essential to students' acquisition of deep scientific understandings and inquiry skills. A key premise of the work reported here is that science inquiry and literacy practices share important properties that make the integration of literacy and science particularly powerful. Participation in investigation-oriented science relies on sophisticated literacy skills, such as the ability to access scientific terminology, interpret arrays of data, comprehend scientific texts, engage in interpretive and critical reading, and read and write scientific explanations (Conley, 2008; Norris & Phillips, 2003; Osborne, 2002). Science provides a setting in which students may be intellectually obligated to set purposes, ask questions, clarify ambiguities, draw inferences from incomplete evidence, and make evidence-based arguments—the very dispositions required as good readers and writers (Cervetti, Pearson, Bravo, & Barber, 2006; Osborne, 2010; Pearson, Moje, & Greenleaf, 2010). When literacy instruction is conceptualized as a form of inquiry, reading and writing can be used to advance scientific inquiry, and students simultaneously learn how to read and how

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

to do science (Cervetti et al., 2006; Greenleaf, Brown & Litman, 2004; McMahon & McCormack, 1998; Pearson et al., 2010).

Despite the potential for science classrooms to contribute opportunities for students to acquire greater proficiency in literacy and science, after decades of science reform efforts, science classes persist in offering few opportunities for students to engage intellectually with either science content or texts (Duschl, Schweingruber, & Shouse, 2006). In one national study of math and science classrooms, nearly two thirds of observed lessons were ranked as low quality overall (Weiss, Pasley, Smith, Banilower, & Heck, 2003). Among the weakest elements of lessons were opportunities for sense-making, use of teacher questioning to enhance understanding, and evidence of intellectual rigor. While lessons were not explicitly rated on key indicators of literacy, assignments involving textbooks were a pervasive feature of science instruction. Furthermore, teachers reported that textbooks influenced both content selection and instructional strategies.

Given the prevalence of reading activities and the influence of textbook programs on lessons that generally fail to engage students intellectually or provide opportunities for sense-making, aligning reading instruction with rigorous academic standards in science can potentially have a significant impact on both students' science and literacy learning. However, while a good deal of recent research points to important elements of successful work in the area of adolescent literacy development (e.g., Kamil et al., 2008; Rycik & Irvin, 2001), there are very few studies of subject-area classrooms, particularly, high school science classrooms, in which these effective features are enacted (cf. Calfee & Miller, 2004; Hart & Lee, 2003; Moje et al., 2004; Romance & Vitale, 2001, 2008, re: elementary or middle school science; Greenleaf et al., 2004, re: high school). Even more rare is an attempt to link increases in literacy achievement to measures of subject-area learning, particularly at the high school level.

In the study reported here, we investigated the relationship between growth in reading achievement and biology learning in the context of high school biology classrooms integrating reading instruction into science teaching. By utilizing an inquiry approach to reading instruction that draws on potential synergies between science and literacy practices, this research attempted to address the question of how to integrate literacy instruction strategically to support complex subject-area learning.

In addition to addressing the challenge of developing students' literacy proficiencies in the context of science learning, this article also advances the idea that we must find effective means to develop the instructional capacity of secondary teachers if we are to succeed in advancing student capacities in science and literacy. Despite the recognized and widespread need for adolescent literacy development in the upper grade levels, very few classrooms provide the needed academic literacy instruction, particularly in the subject areas where it is most critically absent (CCAAL, 2010; C. D. Lee & Spratley, 2010). In large part, high school teachers lack the know-how to

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

simultaneously build students' academic literacy skills and engage them in a rigorous curriculum of subject-area study (Heller & Greenleaf, 2007; Shanahan & Shanahan, 2008). In lieu of helping students develop the proficiencies needed to read, write, and reason with the language, texts, and dispositions of science, typical instructional strategies for struggling readers involve simplifying, slowing the pace, and often abandoning more rigorous course work, virtually assuring low levels of achievement for students who are already behind (Dweck & Molden, 2005; Kamil et al., 2008; Pearson et al., 2010).

National policy responses to persistent achievement inequities have recently targeted the improvement of teacher quality by linking federal funding to state and district reform strategies (for example, Race to the Top, Investing in Innovation, and School Improvement grants), with policies focused primarily on recruiting and retaining better-prepared teachers and distributing these human resources more equitably to schools serving low-achieving student populations (www.ed.gov). Yet teaching is a complex human endeavor, enacted with specific learners in interaction with specific content, materials, and tasks, and decades of research have demonstrated that effective teaching requires a capacity and disposition to learn through reflection while fully immersed in teaching (Ball & Cohen, 1999; CCAAL, 2010; Darling-Hammond & Sykes, 2003; Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Shulman, 1986). In addition to recruiting, retaining, and distributing better-prepared teachers, advancing the academic literacy of all students will depend on identifying and providing effective means of building teacher quality through ongoing teacher professional development. Little systematic research has examined the effects of professional development on teacher practice and student learning outcomes in science (Garet, Porter, Desimone, Birman, & Yoon, 2001; Kennedy, 1998; Killion, 1998; Loucks-Horsley & Matsumoto, 1999), and those studies that exist have been done primarily in the elementary or middle school setting (Calfee & Miller, 2004; Fishman, Marx, Best, & Tal, 2003; Romance & Vitale, 2001, 2006). Further, professional development endeavors vary widely in their structures, formats, and aims (Borko, 2004; Desimone, 2009; Guskey, 2002).

As some researchers have pointed out, professional development opportunities are distributed inequitably, with teachers serving the most vulnerable populations of students receiving training to implement narrow instructional strategies, follow pacing guides, or implement curricula to fidelity rather than professional development opportunities that help them build the flexible repertoires of practice they will need to develop the advanced academic proficiencies of a broader range of the students they serve (Ball & Cohen, 1999; Hargreaves, 2003; Sparks, 2004). As others have argued, professional development is key to standards-based reform (Darling-Hammond et al., 2009; Fishman et al., 2003). Yet in an era in which higher academic standards are envisioned for all learners, the narrow training regimens and implementation monitoring now prevalent in high-poverty schools will fall far short of

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

preparing teachers to meet the lofty instructional goals put forward in such documents as the new Common Core Standards (National Council of Chief State School Officers & National Governors Association, 2010).

In contrast to more narrow conceptions of professional development, the Reading Apprenticeship model of professional development we report on here aims to build teachers' generative capacity to engage with students in collaborative meaning making and problem solving during ongoing instruction (Greenleaf et al., 2004; Greenleaf & Schoenbach, 2004). The study focuses on the impact of professional development designed to foster science teachers' adaptive expertise (Bransford, Derry, Berliner, & Hammerness, 2005; Lai, McNaughton, Amituanai-Toloa, Turner, & Hsiao, 2009), building teacher capacity to surface and model effective ways to address comprehension problems that arise as the varied learners in the classroom interact with course materials, such as biology textbooks, graphs and diagrams, and lab procedures. This study therefore has the potential to advance our knowledge of the role such a model of professional development can play in developing existing teacher quality as well as the potential of such instructional improvement to achieve educationally meaningful increases in student learning outcomes at the high school level. The research base on these vital issues is woefully inadequate to inform policy and practice in educational reform (Desimone, 2009; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007).

Purpose of the Study

This study was designed to examine the effects of an instructional framework, Reading Apprenticeship, and an accompanying professional development model on teachers' ability to integrate disciplinary literacy practices into science teaching in high school biology classes, exploring the resulting changes in teacher knowledge and instructional practices and student achievement in science and reading. This study was a unique collaboration between research and development staff responsible for science programs and literacy programs at a nonprofit educational service agency and independent assessment and evaluation researchers to address critical educational issues related to science and literacy for underachieving high school students. The study targeted schools serving high numbers of African American and Latino students as well as students from low socioeconomic groups and investigated the extent to which literacy integration strategies may differentially impact various subgroups. The research team focused the study on high school biology because it is the science course most high school students are required to take. The study utilized a true group-randomized experimental design and multiple measures of both teacher implementation and student learning over multiple time points to gauge program impacts. Data sources and multilevel analytical methodologies enabled the research team to trace the linkages between a generative model of professional development in

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

literacy for science teachers, science teachers' instructional practices, and student engagement and learning in literacy and science.

Theoretical Background and Relevance to the Field

Literacy Proficiency as a Gatekeeper to the Sciences

Numerous studies have demonstrated that the literacy proficiency of young adolescents shapes their academic futures through systemic inequities that result in tracking students into college-bound and non-college-bound courses of study at the high school level (Hull & Rose, 1989; Knapp, 1995; Oakes, 2005; Sizer, 1992). As early as third and fourth grade, relative success or failure reading subject-area texts begins to shape students' reading engagement and academic achievement (Stanovich, 1986), and differences in reading volume translate to differences in knowledge and vocabulary (Cunningham & Stanovich, 1998). As students move up the grades, continued difficulty comprehending academic texts can shape their choices of courses and their engagement in school (Allington, 1991; Davidson & Koppenhaver, 1993; Guthrie & Greaney, 1991; Guthrie, McGough, Bennett, & Rice, 1996; Guthrie, Schafer, & Hutchinson, 1991), yet students' ability to handle text complexity is the best predictor of success in postsecondary education (ACT, 2006). Furthermore, students' learning outcomes are often measured through standardized achievement tests that require specific subject-area knowledge as well as skillful reading and comprehension abilities. Students' reading proficiency thus becomes a gatekeeper to their further learning in all academic subjects. Improving their capacity to read and comprehend science texts may contribute in important ways to narrowing the achievement gap in science course taking, learning, and achievement.

Academic Literacy: Discipline-Specific Thinking

Over the past several decades, reading has come to be understood as much more than a collection of basic skills. Rather, all texts are shaped by specific conventions and structures of language, and proficient reading of such texts demands the use of these conventions to navigate layers of meaning (e.g., New London Group, 1996; Scribner & Cole, 1981; Shanahan & Shanahan, 2008). Literacy practices become increasingly specialized throughout the school career, reflecting the broader activities that characterize the academic disciplines (Heller & Greenleaf, 2007; C. D. Lee & Spratley, 2010). While reading in all academic disciplines requires advanced literacy skills such as critical thinking, hypothesis testing, effective oral and written communication, and reading across multiple texts and technologies, norms of evidence, logic, and discourse vary widely across content areas (CCAAL, 2010). In science, texts are artifacts of past investigations and are used for reasoning about scientific phenomena. Scientists use texts to generate new research questions and to provide the background necessary for research

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

design and investigation, and skillful reading of science texts mirrors the kinds of thinking characteristic of science exploration and reasoning (Baker, 1991; Greenleaf et al., 2004; Hynd, 1998; Pearson et al., 2010; Roth, 1991; Saul, 2004). The language and texts used to represent and communicate ideas in science present novel challenges of comprehension and interpretation to the science learner (Digisi & Willett, 1995; C. D. Lee & Spratley, 2010; Lemke, 1990; Osborne, 2010; Scott, 1993; Snow, 2010; Their & Davis, 2002; van den Broek, 2010).

Moreover, it is now widely recognized that even skillful reading at early grade levels will not automatically translate into higher-level academic literacy (CCAAL , 2010; Greenleaf, Schoenbach, Cziko, & Mueller, 2001; Heller & Greenleaf, 2007; C. D. Lee & Spratley, 2010; Snow, 2002). As students move up the grade levels, text complexity increases and the uses and features of texts vary from subject to subject. As Paris (2005) has argued, in contrast with decoding, which requires mastery of a small set of constrained skills, acquiring proficiency in reading comprehension requires attainment of a growing set of relatively unconstrained skills that are increasingly situated in particular texts and reading tasks. Paris's distinction calls for developing reading instruction that is commensurate with the authentic literacy tasks characteristic of advanced academic learning. All considered, the implications of these challenges for the literacy and science learning of diverse populations of students are profound. Increasingly, students in U.S. schools come from a variety of economic, linguistic, cultural, and ethnic backgrounds, bringing significantly different experiences and expectations about how to initiate and sustain conversations, how to interact with teachers and peers, how to identify and solve different types of problems, and how to go about particular reading and writing tasks (Greenleaf, Hull, & Reilly, 1994; C. D. Lee, 1995; Moje, Dillon, & O'Brien, 2000). Literacy researchers have therefore argued that for all students to learn to perform academic literacy tasks, teachers need to make explicit the tacit reasoning processes, strategies, and discourse rules that shape successful readers' and writers' work in particular disciplines (e.g., Delpit, 1995; Fielding & Pearson, 1994; Freedman, Flower, Hull, & Hayes, 1995; Gee, 1999; Lemke, 2006; Moje, 2008; Pearson et al., 2010; Shanahan & Shanahan, 2008). These researchers have advocated explicitly showing students how to carry out literacy tasks, building bridges from their cultural knowledge and language experiences to the language and literacy practices valued and measured in school in particular academic subject areas.

Integrating Literacy Apprenticeships Into Subject-Area Teaching

There is general agreement that direct and explicit comprehension instruction is essential to effective adolescent literacy instruction (CCAAL, 2010). Yet recent research has identified problems with comprehension instruction as it is currently implemented in many content-area classrooms. Conley (2008) has questioned the pervasive practice of applying generic

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

strategies designed for young children to the teaching of complex content area goals and materials. In science, where a single text may communicate misconceptions or provide an incomplete picture, getting the gist of a text may be only the first stage in a more sophisticated process of questioning and deliberation. To be proficient readers of science, students need discipline-based reading strategies that permit them to go considerably beyond literal comprehension characteristic of early literacy tasks to tackle academic tasks, such as reconciling multiple texts with different methodologies, perspectives, interpretations, and biases (CCAAL, 2010; Conley, 2008; Kamil et al., 2008). Other researchers have likewise underscored the urgency of creating conceptually rich discipline-based skills instruction aligned with science learning goals to challenge students intellectually while helping them build their skills in high-level literacy (Paris, 2005; Schoenbach & Greenleaf, 2009; Umphrey, 2009).

Recent critiques also raise concerns that comprehension strategies instruction can potentially displace attention to learning in the subject areas, becoming an end in itself (Conley, 2008; Kamil et al., 2008). Contrasting the use of graphic organizers in two science lessons, Conley (2008) argues that "there is a significant but overlooked difference between using cognitive strategies as a 'teaching tool' versus using cognitive strategies as a 'learning tool'" (p. 87) and challenges the assumption that teaching a cognitive strategy through activities such as having students contribute details to teacher-generated graphic organizers will help students learn to organize their own thinking. Instead, he suggests that well-integrated cognitive strategy instruction should function "as a deliberate action to develop in students a critical understanding of subject matter ideas *and* a cognitive approach to learning" (Conley, 2008, p. 91).

Various studies over the past few decades have demonstrated the value of integrating the explicit teaching of comprehension, text structures, and word-level strategies into compelling sense-making activities with texts for student reading achievement (Baumann & Duffy, 1997; Beck, McKeown, Hamilton, & Kucan, 1997; Guthrie et al., 1996; Kamil et al., 2008; Pressley, 1998). A recent study of reading and writing about science at the intermediate level indicates that when upper elementary students are explicitly taught strategies for science reading and writing in a learning environment structured to support collaboration and metacognition, students' reading and writing of science content improves (Miller, 2004). The authors of this study argue that literacy instruction is best when embedded in meaningful content instruction (Calfee & Miller, 2004). At the elementary level, a curriculum designed to integrate literacy and science learning is showing promise in efficacy studies (Cervetti et al., 2006). Similarly, recent reports of the National Reading Panel (2000), Institute of Education Sciences (IES; Kamil et al., 2008) and CCAAL (2010) have all concluded that teaching a combination of reading comprehension techniques, rather than teaching individual comprehension strategies in isolation from one another and from content instruction, is

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

the most effective method to increase reading comprehension. Furthermore, these reports support the executive role of metacognition in using strategies effectively. For example, CCAAL (2010) recommends, "Once strategies are introduced, students must also learn how to think metacognitively, that is, to determine which strategy is appropriate for a given reading task" (p. 77). While the authors of this report identify English learners as particular beneficiaries of metacognition, the significant vocabulary load of science texts (Groves, 1995) suggests that the ability to marshal reading strategies to compensate for the comprehension-inhibiting effects of unfamiliar vocabulary and concepts is an essential component of effective literacy instruction in high school science. Student collaboration and instructionally focused conversation have also been identified as key to improving literacy achievement in major adolescent literacy policy initiatives, including the CCAAL report, the IES practice guide (Kamil et al., 2008), and the report of the National Reading Panel as well as the RAND (Snow, 2002) report. A recent large-scale study of science instruction suggests that such collaborative meaning-making activities are rare in high school science classrooms (Weiss et al., 2003).

In order to develop in students the metacognitive approach to learning required for high-level literacy, research on effective comprehension instruction thus calls for a kind of flexibly adaptive teaching that is neither commonplace nor simple (Heller & Greenleaf, 2007). Some have adopted the metaphor of "cognitive apprenticeship" to describe teaching designed to assist students in acquiring more expert, or proficient, cognitive processes for particular valued tasks, such as reading comprehension, composing, and mathematical problem solving (e.g., Bayer, 1990; Brown, Collins, & Newman, 1989; Lave & Wenger, 1991). When the target proficiency is a cognitive practice, such as composing or comprehending a text, the invisible mental processes involved in the task must be made visible and available to apprentices as they actually engage in meaningful literacy activities (Donovan, Bransford & Pellegrino, 1999; Pearson, 1996; Freedman et al., 1995). To help students develop as readers and writers, teachers can create "literacy apprenticeships," engaging students in meaningful and complex literacy practices while demystifying and supporting students in practicing these literacy practices (Brown et al., 1989; C. D. Lee, 1995; Osborne, 2002).

The Reading Apprenticeship Approach to Integrating Science and Reading Instruction

Based on this research in literacy learning, to support teachers' learning and adolescents' discipline-specific literacy development, Greenleaf, Schoenbach, and their colleagues have developed, implemented, and studied the impact of an instructional model for academic reading instruction: the Reading Apprenticeship instructional framework (Greenleaf et al., 2001; Schoenbach, Greenleaf, Cziko, & Hurwitz, 1999). While the framework incorporates research-based instructional approaches that have been shown

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

to improve adolescent literacy levels— including vocabulary and academic language development techniques, direct and explicit comprehension strategy instruction, extended discussion of text meaning and interpretation, and strategies to increase student motivation and engagement in literacy learning (Kamil et al., 2008)—in this model, reading instruction is also closely aligned with subject-area learning goals, in this case, science, and integrated into content-area teaching, rather than being an instructional add-on or additional curriculum. Further, unlike approaches that respond to the challenges of science reading by simplifying texts or tasks (Pearson et al., 2010), the Reading Apprenticeship framework helps teachers to improve their instructional routines around existing science curriculum and texts and to extend students' reading opportunities with these and ancillary science materials.

The Reading Apprenticeship model is based on research indicating that most students are capable of complex thinking and of carrying out scientific and literary inquiry but have not been given the skills or self-confidence to approach these tasks effectively (Greenleaf et al., 2001; Langer, 2001; C. D. Lee & Spratley, 2010; Moje, Overby, Tysvaer, & Morris, 2008). In this framework, students are given extended opportunities to read a wide range of texts with instructional support—both textbook and lab materials and ancillary materials, such as journal articles and trade books. Through an "apprenticeship" process, content-area teachers explicitly model and guide students in practicing the tacit reasoning processes, strategies, and discourse rules that shape successful readers' and writers' work. The focus of instruction is therefore on engaging with science texts to actively construct meaning and flexibly utilizing an array of comprehension tools in order to do so rather than on learning to carry out a particular comprehension strategy or set of strategies as learning targets in their own right (Conley, 2008; Kamil et al., 2008).

The Reading Apprenticeship framework centers on metacognitive conversation, involving explicit metacognitive routines, modeling, small-group work, and class discussions that focus, in the case of science instruction, on how to read science and why people read science materials in the ways they do as well as on the science content of what is read in science classes. These discourse routines offer students support to clarify content, discuss the processes they use in reading and problem solving, practice comprehension strategies, respond to and elaborate on content, engage in word-learning strategies, write to learn and to consolidate learning, and make connections to other related texts and topics. Reading Apprenticeship practices are designed to draw both on what teachers know and do as readers in particular academic domains and on adolescents' underestimated strengths as learners. Many of the underperforming students in U.S. high schools have resigned themselves to low literacy and academic attainment (C. D. Lee & Spratley, 2010). In this framework, teachers attend to students' affective and identity issues, creating relevant and affectively safe learning opportunities that help students build stamina and dispositions to engage in academic

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

tasks, discipline-based literacy practices, and inquiry and to develop identities as resilient learners (Schoenbach & Greenleaf, 2009). The framework thus aims to support teachers in building students' capacities to carry out close, intellectually engaged reading; gain insight into their own thinking processes; make meaning; acquire academic and disciplinary language; read independently; and set personal goals for literacy development.

Previous studies of the impact of Reading Apprenticeship have demonstrated increased reading achievement and academic engagement across a diverse group of adolescents enrolled in a Reading Apprenticeship Academic Literacy course in ninth grade (Greenleaf et al., 2001; Schoenbach et al., 1999). These results have been replicated in additional studies, including a randomized, controlled study examining whether supplemental literacy classes improve the reading skills of struggling ninth-grade readers (Somers et al., 2010), demonstrating that teachers' implementation of Reading Apprenticeship can result in significant gains for students across varied grade levels and subject areas (for a review of these studies, see www.wested.org/ra). Further, explicit support for reading in a chemistry class has been shown to build low-performing students' abilities and dispositions to work through conceptually dense science materials and, ultimately, to participate in science learning in new ways (Greenleaf et al., 2004; Litman & Greenleaf, 2008). This prior research suggests that implementation of the Reading Apprenticeship instructional framework has the potential to increase students' reading and biology engagement and achievement at the high school level.

An inquiry model of science and literacy integration. The vision of science and literacy integration proposed in this model contrasts sharply with current instructional practices in science classes nationwide. In science education reform, the reading of science texts is often contrasted to opportunities for students to build science knowledge through hands-on scientific exploration (e.g., Bybee, 1995, 1997; National Committee on Science Education Standards and Assessment, 1996; National Research Council, 1990). Many science teachers are keenly aware of students' difficulty comprehending science texts and are not certain how to integrate science reading experiences with hands-on investigations. As a result, in recent years, there has been a widespread reduction of reading in secondary science classrooms, precisely as policymakers are raising alarms about the reading proficiencies of adolescents. While investigations are promoted in science education reform because they convey the nature of science as a discipline and help students to acquire science practices (American Association for the Advancement of Science, 1993), in practice, hands-on activities can serve a less noble purpose by helping teachers avoid the problem of students' lack of academic literacy proficiency. Furthermore, without the benefit of textual support, hands-on instruction may degrade to activity for activity's sake, with students expected to infer the important science ideas underlying their problem solving and laboratory investigations on their own, or teachers may

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

short-circuit student learning by telling students what they should have found in their investigations (Weiss et al., 2003).

As a high school life science teacher participating in a large-scale study of science instruction told researchers,

> With that particular class, because a lot of them in there are either slow readers, or nonreaders or remedial readers, they don't get anything out of reading stuff out of the textbook. So they actually have to be doing . . . They have to do, so they like to do a lot of hands-on stuff. (Weiss et al., 2003, p. 96)

The following description, drawn from a random sample of typical 9th- to 12th-grade science lessons, illustrates approaches to reading in science, and science instruction in general, commonly found in that study:

> This was the final [9th- to 10th-grade biology] lesson in a unit on evolution. . . . The teacher began the lesson with a review, asking a series of questions such as: "What was the first piece of evidence for evolution? Who remembers what epoch we live in? What was Lamarck's theory of evolution? Why did Darwin say [giraffes'] necks get long?" Next, the teacher told students to open their texts to the appropriate page, and placed an outline on the overhead projector. For the next thirty minutes she lectured based on that outline. She gave five modern pieces of evidence for evolution by natural selection: analogous structures, homologous structures, vestigial organs, similarities of embryos, and macromolecules. She provided examples for each (from the textbook) and asked questions throughout, such as "What do you notice about these diagrams (of arm bones of different mammals)?" "Can anyone think of an organ inside the body that today is not used?" and "What does a nucleotide consist of?" When directly called to do so, students answered questions, but otherwise spent most of their time copying the teacher's outline into their notes. The teacher next turned her lecture to the subject of evolutionary patterns listing three types: coevolution, convergent evolution, and divergent evolution. She drew diagrams and gave examples for each. (Weiss et al., 2003, Appendix H-5)

This description of what Weiss and colleagues found to be typical of science instruction presents the teacher as conveyor of information to students, whose role is limited to dutifully noting the information they are served and retrieving it on demand, and reading serves as precursor to comprehension testing rather than an authentic role in science inquiry and learning. The following lesson from field notes taken during an observation in a Reading Apprenticeship biology classroom involves similar content to the lesson described in the Weiss et al. (2003) study, yet illustrates significantly different roles for learners, for teachers, and for science reading:

> The lesson was a department-mandated benchmark assignment that involved reading a dense four-page text, "Island Biogeography and

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

> Evolution: Solving a Phylogenetic Puzzle with Molecular Genetics" (http://www.ucmp.berkeley.edu/fosrec/Filson.html). The teacher noted that the introductory paragraph was extremely dense with terminology and that the best way to get started was to pull out the evolution terms that they already knew to make sure these were clear. Saying, "Okay, let's figure this out," she had students read the paragraph silently and identify all the terms that they knew are related to evolution. Students then worked with partners to define the terms. Each pair shared one term with the whole class for a quick discussion review. The teacher then directed students' attention to a diagram of lizards on page 2 and asked for a few thoughts about what they noticed. Observations included the fact that lizard and island sizes corresponded (i.e., small lizards with small islands). After looking at the diagram, the teacher had students return to page 1 and read the next two paragraphs silently, saying it would tell them something about the lizards. She also pointed out the graphic and reminded them to use the picture to help them figure out the meaning. After reading individually, students worked with partners to generate at least two questions about the reading to be discussed the following day. The teacher reminded students to ask questions about what they didn't understand, as well as what they wanted to learn more about. Students recorded their questions on whiteboards. Questions included such things as, "How did the lizards get on the islands?" "Why are the lizards different sizes on different islands?" "What formed the islands over the last 20 million years?" (R. Rosen, field notes, March 15, 2007)

In this class, problem solving with complex science texts serves the goal of science learning, and students are actively engaged in the intellectual work of inquiry. The teacher orchestrates a metacognitive conversation that draws students' confusions and contributions into the whole-class discussion as resources for problem solving, invests time in setting up tasks to engage collaborative groups of students in profitable work, and invites students' curiosity and questions dynamically to drive the learning process. This responsive mode of instruction requires teachers to navigate interactions between learners, texts, tasks, and science content (Ball & Cohen, 1999). The problem facing science education reform, and literacy research more generally, is how to transform traditional instruction in which teachers convey and test students' retention of information about science into the kind of active, inquiry-based, apprenticeship learning opportunity evident in this classroom. While attempts to address this problem have included developing new curricula to model active modes of learning that integrate reading into science learning; rewriting science texts to provide more easily accessible science concepts to learners; reforming teacher education programs to integrate science departments, schools of education, and literacy methods; and promoting lesson routines as protocols for instruction (Pearson et al., 2010), investing in professional development must also play a key role in building science teachers' capacity for responsive teaching consonant with the inquiry practices of science (Loucks-Horsley, Love, Styles, Mundry, & Hewson, 2003;

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Romance & Vitali, 2001, 2008; Supovitz, Mayer, & Kahle, 2000). Further, to prepare teachers to support active student inquiry into science texts and investigations, the professional development must aim for knowledge transformation and transfer to new problems that cannot be solved by previous routines rather than focus on training teachers to implement a more constrained set of specific procedures (Bransford et al., 2005; Lai et al., 2009; Supovitz & Turner, 2000).

The Knowledge Base on Effective Methods of Professional Development

If reforming science education toward inquiry learning has been a challenge, the project of integrating literacy instruction into content-area instruction has been equally daunting. A long history of research in reading has demonstrated that reading comprehension strategies are not often taught in subject-area classes, even when teachers are trained to use these strategies during subject-area teaching (Alvermann & Moore, 1991; Duffy et al., 1986; Duke, 2000; Durkin, 1978; Fielding & Pearson, 1994; Richardson, 1994; Snow, 2002). Furthermore, even when teachers do implement literacy strategies, they often have difficulty balancing content and strategy instruction at the same time because a culture of whole-class direct instruction makes it difficult for teachers to engage students actively in reading and learning (Reed, 2009). To move in the direction of inquiry-oriented science instruction and to increase the potential for reading instruction to become tightly integrated into biology teaching, professional development must therefore demonstrate features of high-quality learning for teachers that are known to be effective in producing changes in classroom instruction (Kamil et al., 2008; Porter, Garet, Desimone, & Birman, 2003; Strickland & Kamil, 2004; Wei et al., 2009).

There is much known and much unknown about the elements of effective professional development. In 1999, the National Research Council report on the science of learning identified important themes in teacher learning (Bransford, Brown, & Cocking, 1999). Other studies examining the impact of teacher professional development on teacher knowledge and practice in both science and literacy have supported and elaborated these findings (Garet et al., 2001; Kennedy, 1998; Loucks-Horsley & Matsumoto, 1999; Reed, 2009; Yoon et al., 2007). Desimone (2009) argued recently that there is now sufficient empirical evidence to establish consensus on a set of five core features of effective professional development: a content focus on student learning in particular subjects, active learning opportunities, coherence with teachers' existing practices and policy contexts, sufficient duration, and collective participation. Similar elements appear in various guises in a wide range of research on teacher professional development (Ball & Cohen, 1999; Darling-Hammond et al., 2009; Garet et al., 2001; Guskey & Huberman, 1996; Kennedy, 1998; Klingner, 2004; Little, 2001; Shulman, 1987). Yet surveys of professional development surface large discrepancies between what is

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

known to be effective and teachers' professional development experiences (Fishman et al., 2003; Loucks-Horsley & Matsumoto, 1999; Richardson, 2003).

While there is a growing consensus around what constitutes effective professional development, the ambitious research agenda inspired by the National Research Council report on the science of learning made only passing mention of the need for studies linking teacher professional development, teacher learning, and student achievement (Donovan et al., 1999). As Desimone (2009) has outlined, early research on teacher professional development established effectiveness through teacher self-report on such outcomes as teachers' satisfaction with the staff development experience, and later studies measured changes in participants' beliefs and attitudes, content knowledge, or commitment to change. Some research has also measured changes in participants' practice. Relatively few studies of professional development programs have examined the impact of teacher learning on student performance (Desimone, 2009; Killion, 1998). Of 450 professional development projects identified and reviewed by the Middle Grades Initiative of the National Staff Development Council, fewer than 10% included any measurement of student achievement (Killion, 1998). Likewise, a review of professional development in math and science identified only four science programs that collected data on student achievement (Kennedy, 1998). Thus, while research has confirmed that effective professional development can increase teacher confidence in instructing students with diverse abilities (Garet et al., 2001), few studies can link professional development to student achievement (Garet et al., 2001; Kennedy, 1998; Loucks-Horsley & Matsumoto, 1999; Reed, 2009). A decade later, there is still limited evidence to confirm that professional development can increase students' learning in the content areas (Desimone, 2009; Reed, 2009; Yoon et al., 2007).

Compounding the lack of attention to student outcome data in many studies of professional development is the methodological challenge of demonstrating an impact of teacher professional development on student achievement (Borko, 2004; Loucks-Horsley & Matsumoto, 1999; Reed, 2009; Supovitz, 2001; Yoon et al., 2007). A recent review of the effects of teacher professional development on middle school teachers' subsequent implementation of literacy strategies in content-area classes yielded only one study that included student outcome data, and that study did not include statistical information suitable for confirming the calculation of effect sizes (Reed, 2009). The most comprehensive review of the impact of teacher professional development on student achievement in science, math, reading, and English or language arts identified 9 studies out of more than 1,300 potential studies that met criterion standards of evidence (Guskey & Yoon, 2009). No studies at the middle or high school levels met these standards, nor did any study published after 2004. The authors of the review acknowledged that they were stunned by these findings: "Obviously, these findings paint a dismal picture of our knowledge about the relationship between professional development and improvements in student learning" (Guskey & Yoon, 2009, p. 497).

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Elements of Effective Professional Development

To improve the quality of inquiry into teacher learning toward more rigorous standards of evidence, Desimone (2009) proposed the use of a set of core features and a common conceptual framework for professional development impact studies, identifying five characteristics of professional development that have been shown to be critical to improving teachers' knowledge, skills, and classroom practices and thus may hold promise for increasing student achievement. As mentioned earlier, these elements include (a) content focus, (b) active learning, (c) coherence, (d) duration, and (e) collective participation. We briefly review some of the relevant research below in order to describe the design elements of Reading Apprenticeship professional development within this common conceptual framework.

Content focus. While early research on effective professional development focused primarily on structural variations among programs, such as the length of in-service, the distribution of sessions over time, whether participants were volunteers or were mandated to participate by school or district policies, and whether teachers attended in groups or as individuals, subsequent studies of the impact of professional development have determined that the benefits of a strong content focus outweigh the influence of structural features (Kennedy, 1998; Loucks-Horsley & Matsumoto, 1999; Reed, 2009; Yoon et al., 2007; Zech, Gause-Vega, Bray, Secules, & Goldman, 2000). In particular, professional development approaches that focus on developing pedagogical content knowledge through such means as adapting curriculum materials, analyzing student thinking, examining cases of teaching and learning, and reflecting on and critiquing one's own and others' teaching can develop teachers' pedagogical content knowledge (Shulman, 1986, 1987) and have a positive impact on student science and literacy learning (Abell, 2008; Guskey & Yoon, 2009; Kennedy, 1998; Loucks-Horsley & Matsumoto, 1999; Reed, 2009; vanDriel, Beijaard, & Verloop, 2001). In summarizing findings on the impact of professional development on student achievement in math and science, Kennedy (1998) noted

> In the studies reviewed here, programs whose content focused mainly on teachers' behaviors demonstrated smaller influences on student learning than did programs whose content focused on teachers' knowledge of the subject, on the curriculum, or on how students learn the subject. Moreover, the knowledge that these more successful programs provided tended ***not*** to be purely about the subject matter—that is, they were not courses in mathematics—but instead were about ***how students learn*** that subject matter. (p. 17)

Thus, Kennedy found offering teachers specific teaching techniques and prescriptions for structuring lessons to be less effective than focusing on student learning and problem solving, noting that the modeling techniques

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

presented in these programs reflected a "cognitive apprenticeship" approach to teaching and provided evidence for the positive impact of such an approach on student learning.

A recent comprehensive review of research on the impact of teacher professional development on student achievement in three content areas likewise found that where professional development increased student achievement, professional development invariably centered directly on helping teachers better understand both what they teach and how students acquire specific content knowledge and skill (Guskey & Yoon, 2009). In order to navigate the complexities of teaching described earlier, teachers will need professional knowledge that is generative, that is, knowledge based in deep understandings of how students learn particular subject matter that will allow them to generate effective responses to students' thinking in the moment of teaching (Boston & Smith, 2009; Cohen & Ball, 2000; Schoenbach & Greenleaf, 2009). Thus an inquiry orientation is likely to be more effective than professional development focused on teacher behavior alone (Kennedy, 1998; Loucks-Horsley & Matsumoto, 1999).

Duration. While research has established the primacy of content over structural features of professional development, intellectual and pedagogical change takes time, and research indicates that professional development activities must be of sufficient duration for that change to occur (Darling-Hammond et al., 2009; Donovan et al., 1999). In addition to enough contact hours, virtually all studies that showed positive outcomes in student learning in a comprehensive review of the impact of professional development on student achievement included structured and sustained follow-up after the main professional development activities (Yoon et al., 2007), and larger effect sizes on student achievement in science have been found when professional development is distributed across the school year (Kennedy, 1998).

Yet this does not necessarily mean that improving student achievement requires ongoing school-based reform. While the nine studies that met evidence standards in the Yoon et al. (2007) review varied in duration and intensity, every study that showed a positive impact on student learning took the form of workshops or summer institutes—albeit workshops that engaged teachers in active learning and provided them with opportunities to adapt the practices to their unique classroom situations. Furthermore, significant student learning has resulted from professional development with as little as 16 hours total contact distributed over five sessions (Kennedy, 1998). No successful program reviewed by Yoon et al. used a school-based professional learning approach, such as peer coaching or collaborative problem solving, and schoolwide professional development produced the smallest effect sizes in student math and science achievement (Kennedy, 1998). Lack of variability in the form of professional development precluded Yoon et al. from determining the relative merits of school-based or workshop-based professional development. However, research suggests that effective professional development can be accommodated by well-designed workshops and

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

institutes. This finding has great practical import, since teachers' available time for professional development is a scarce resource in educational improvement. Given the many demands on teachers' time and attention, those responsible for professional development have the obligation to work as strategically and efficiently as possible in developing teacher capacities by designing effective learning environments for teachers (Greenleaf & Schoenbach, 2004).

Collective participation. Research into teacher learning has shown that encouraging discourse around problems of practice helps teachers to delve deeply into issues of teaching and learning (Borko, 2004; Fullan, 1991; Guskey, 2002; Loucks-Horsley & Matsumoto, 1999; Richardson, 1994). Collective participation of teachers from the same school, grade, or department provides an opportunity for interaction within professional learning communities (Camburn, Rowan, & Taylor, 2003; Clark & Borko, 2004; Grossman, Wineburg, & Woolworth, 2001; McLaughlin & Talbert, 2001). Even when such collective participation from school sites may not be possible, elements of professional development can stimulate interaction and discourse among participants around problems related to teaching—for example, by examining and analyzing artifacts of practice, such as video cases and student work, for evidence of student thinking and teacher decision making (Desimone, Garet, Birman, Porter, Yoon, 2002; Fennema et al., 1996; Kazemi & Franke, 2004; Little, 2001, 2002; Shulman, 1986, 1987; vanDriel et al., 2001).

Active learning. Active modes of sense-making and inquiry into teaching and learning have long been related to effectiveness of professional development (Ball & Cohen, 1999; Guskey & Huberman, 1996; Kennedy, 1998; Loucks-Horsley et al., 2003; Sparks & Hirsch, 1997; Wilson & Berne, 1999). Professional development that engages teachers in collective participation and discourse around problems germane to practice and that helps teachers more deeply understand the teaching and learning of particular content will necessarily entail opportunities for teachers to engage in active learning (Borko, 2004; Darling-Hammond et al., 2009; Garet et al., 2001).

Coherence. Finally, Desimone (2009) identified coherence as a core feature of effective professional development. According to Desimone, coherence involves both the extent to which teacher learning is consistent with teachers' existing knowledge and beliefs and the degree to which it promotes coherence between teachers' professional development experiences and their experiences of school, district, and state reforms. In our view, however, there is little research support for the benefits of providing professional development experiences that are entirely consistent with teachers' knowledge and beliefs. To the contrary, Huberman (1995) observed 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. The beliefs teachers bring to professional development

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

exert a powerful influence on their development of new concepts and understandings, and professional development needs to engage these preconceptions at the same time as it challenges them (Pajares, 1992; Spillane, 1999). Yet to transform the nature of instruction, it is essential to change the theories that inform teachers' decision-making processes (Boston & Smith, 2009; Bransford et al., 2005; vanDriel et al., 2001; Whigham, Andre, & Yang, 2000). Rather than providing professional development consonant with teachers' current beliefs and practices, therefore, transforming instruction requires professional development that challenges teachers' beliefs and practices, provides support for resolving the resulting cognitive dissonance, and at the same time responds to teachers' needs for pragmatic solutions to problems of practice by offering teachers a repertoire of tools and strategies consistent with these new understandings (Greenleaf & Schoenbach, 2004; Loucks-Horsley & Matsumoto, 1999).

The second dimension of coherence identified by Desimone (2009), the consistency of school, district, and state reforms with what is taught in professional development, may be an important element of the context that mediates and moderates links between professional development, teacher learning and practice, and student learning rather than a core feature of effective professional development itself. Here, too, the degree of coherence with existing institutional factors may be problematic, as may be the case when professional development based on theory and research collides with state-, district-, or school-level initiatives and mandates. From this perspective, effective professional development is coherent in that teacher learning must be situated in classroom practice, and cognitive dissonance and resolution of dissonance must occur in relation to teachers' own students and context (Lai et al., 2009; Loucks-Horsley & Matsumoto, 1999). For changes in practice to occur, effective professional development must help teachers metacognitively engage their own prior knowledge and conceptions as well as acknowledge the contextual variables that impact their teaching and their students' learning (Phelps, Graham, & Kerr, 2004).

Design Elements of Reading Apprenticeship Professional Development

Based on these understandings of the important features of professional development and with reference to Desimone's (2009) proposed conceptual framework for professional development impact studies, below we describe the elements of the Reading Apprenticeship professional development model that constituted the intervention for this study before turning to the study methods and findings. Based on the Reading Apprenticeship framework and commensurate with the contextualized nature of academic literacy, the professional development model utilized in this study is designed to transform teachers' understanding of their role in adolescent literacy development and build enduring capacity for literacy instruction in the academic disciplines (Greenleaf & Schoenbach, 2004). We outline the specific

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

features of teacher learning opportunities in this model using Desimone's proposed conceptual framework.

Content focus and active learning. Reading Apprenticeship professional development is inquiry based, subject-area focused, collaborative, and designed to address teachers' conceptual understandings as well as practical implementation needs. The professional development model emphasizes the development of pedagogical content knowledge (Shulman, 1986, 1987). From the outset of the professional development, teachers are immersed in learning experiences integrating science and literacy by engaging in rich investigations into science reading and science investigation. The instructional framework positions literacy as inquiry, and professional development activities aim to draw on the similarities between science inquiry processes and literacy. Professional development routines engage teachers in learning experiences integrating science and literacy by engaging in rich investigations into science reading and science investigation.

Because the content focus of this professional development model centers on learning and literacy thinking processes, it is tightly linked to the active mode of learning designed for teachers and for their students. To increase teachers' capacities to generatively design and implement the kind of instruction that supports student literacy, the professional development immerses teachers in models of practice that we aim for them to create in their own classrooms: inquiry-based, collaborative classroom instruction that engages students actively in metacognitive conversations about reading and learning processes. To this end, in professional development sessions, teachers participate in carefully designed inquiries to help them unlock their own disciplinary expertise in relation to literacy. They work to identify the features of disciplinary texts that might present stumbling blocks to learners. Most importantly, they collaboratively investigate student work, videotaped classroom lessons, and case studies of student literacy learning designed to foster new expectations of what their own students can accomplish. In professional development sessions, they practice 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.

The inquiry processes for teacher development are the result of an ongoing research and development process involving varied communities of teachers and researchers over the past decade to iteratively refine effective inquiry designs for professional development (Greenleaf & Schoenbach, 2001, 2004). The resulting collaborative inquiries are designed to engage teachers in exploring (a) metacognitive processes involved in reading complex texts, (b) videotapes of metacognitive process interviews with students reading complex texts, (c) evidence of student thinking in samples of student work, (d) videotapes of classroom lessons in which teachers attempt to integrate Reading Apprenticeship instructional approaches into context

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

learning, (e) the varied types of texts used to represent ideas in the discipline, (f) the knowledge and language demands of disciplinary texts, and (g) the benefits and potential pitfalls of using specific reading strategies with subject-area texts. Elsewhere we have described these tools and approaches in more detail and explicated their specific aims for developing teachers' capacity for responsively adaptive teaching: in brief, to build teachers' conceptions of reading in their disciplines, to build teachers' insights into student learning needs and capacities, and to build teachers' situated and conditional use of reading comprehension strategy instruction (Greenleaf & Schoenbach, 2004).

These inquiries also develop teacher capacity to participate in and facilitate the collaborative metacognitive conversations at the center of the Reading Apprenticeship instructional framework. Metacognitive conversations are framed in social routines that support talk about thinking and reading, such as think-aloud, think-write-pair-share, and reciprocal small-group discussions about written notes or annotations centered on reading processes with science texts. As teachers read, surface, and discuss their problem-solving responses to challenges they find in texts and share their reading processes, the distributed knowledge in the room about how to strategically approach reading becomes shared knowledge through visual note-taking procedures. Professional development facilitators help teachers to label their reading processes and thereby develop declarative knowledge about—that is, a language for describing—reading and thinking processes. Guided practice of reading strategies and discussion of how these strategies support reading comprehension aim to build procedural and conditional knowledge about how and when to monitor comprehension and resolve confusions with science texts. Routines for metacognitive conversation serve as a model for classroom instruction that is designed to support teachers to sustain an ongoing inquiry into reading processes in their classrooms. These inquiry approaches are meant to support teachers' ongoing learning about reading processes as well as to help their students build and use new strategic approaches to comprehension challenges, all while reading (and learning) science.

As an example of one such inquiry, teachers might engage in a phylogenetic investigation involving a finch photo sort. The inquiry begins with a partner think-aloud, guided by a set of disciplinary questions: What are the physical characteristics of each bird? Which birds might be most closely related? What makes you think so? What questions do you have? While one partner makes observations, problem solves, and thinks aloud, the other partner listens and scripts what the first partner is saying. Pairs then form groups of four, and after sharing their most interesting observations, reflections, questions and conclusions, the group will discuss its ideas about which finches might be most closely related. Together, group members construct a claim and support it with evidence from the photographs. Finally, the group designs a whiteboard explanation of its hypothesis about relationships

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

between the finches to share with other groups that includes a schematic representation of how the group thinks the finches are related, observations and evidence that support its claim, relevant background knowledge, connections and interpretations that could be researched and substantiated, questions, research limitations and unresolved issues, and a proposal for additional research.

The pairs and small groups follow this hypothesis construction by conducting similar collaborative inquiries as they read and discuss a set of texts about Darwin's field studies of the Galapagos finches. Teachers use their own experience from this investigation as background for a classroom video case inquiry exploring student thinking and learning during a similar activity. Just as with the finch inquiry, classroom video cases are treated as texts for common inquiry, sense-making, and argumentation, and metacognitive conversation routines and protocols developed for video discussion are designed to support teachers in building knowledge of reading, insights into student thinking, and ideas for effective literacy instruction from the video case and collaborative discussion.

Duration. Reading Apprenticeship professional development is designed to support teacher learning over a 2-year sequence of daylong sessions, with summer institutes followed by follow-up sessions within the school year. Daylong sessions provide teachers with a sufficient stretch of time in which to engage in active inquiry as learners, analyze the pedagogical structure of these opportunities, reflect on the impact of these pedagogies on their learning and that of their students, and plan for instruction. Between professional development sessions, teachers are asked to use new pedagogical tools in their classes and practice new ways of responding to students. They are asked to bring student work resulting from trying out these new instructional approaches to subsequent sessions, and in collaboration with their colleagues, they engage in examination of student work, student thinking processes, and student instructional needs. The aim of the ongoing sessions is to help teachers "diagnose problems in their classrooms and schools, apply evidence-based and often alternative solutions to them and evaluate and analyze the impact of implemented procedures" (Darling-Hammond et al., 2009, p. 29). Through this cycle of inquiry, the aim is for teachers to receive continuing help in surfacing new issues and problems and ways of understanding and translating them into new practices (Loucks-Horsley & Matsumoto, 1999).

Collective participation. Instructionally focused conversation is at the heart of the Reading Apprenticeship instructional framework, and Reading Apprenticeship professional development is likewise organized to promote discourse about literacy, content, and problems of practice. The texts and reading tasks used in professional development sessions are designed to raise authentic problems for teachers. As they discuss how to resolve the reading challenges they experience, teachers are immersed in a model of inquiry-based

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

strategy instruction—a metacognitive conversation about how to identify and resolve comprehension problems. Following this immersion, participants debrief the experiences, making the pedagogy embedded in the activity apparent in order to build knowledge of how to support metacognitive conversation about reading processes in the classroom. Teachers are asked to describe the pedagogies that supported their own reading and learning, their thinking about reading, and collaborative talk about thinking and reading. Teachers are invited to reflect on how metacognitive conversation about reading supported their own learning and to extend the conversation to classroom implications of such learning opportunities and needed adaptations for instruction with students. By understanding the impact of the designed inquiries and their implications for instruction, teachers build an understanding of the inquiry processes and their own purposes for learning how to implement these instructional routines. In this way, teachers work to solve problems of practice in collaboration with their colleagues as they take ownership of these problems and design instruction that supports literacy in science.

Typically, teachers in Reading Apprenticeship professional development participate in school-based teams organized into cross-site learning networks to provide opportunities for schoolwide professional conversation and collective action on literacy in content areas. In the current study, however, constraints of the study design made this organization impracticable, as will be detailed in the study design, below.

Coherence. By design, Reading Apprenticeship professional development activities confront many deeply held beliefs and commonly accepted practices in traditional secondary science education, among them simplistic views of reading, misperceptions about the substantial capabilities of their diverse students, and little appreciation of the roles reading and science texts play in science learning. In order to help teachers build a new set of conceptual understandings and analytical thinking tools for teaching literacy in science and responding productively to students' literacy needs as they read and learn in the science classroom, professional development activities engage teachers in examining video and written case studies of student reading designed to encourage evidentiary thinking and to surface and challenge teachers' current beliefs and practices. As teachers share and discuss their observations and interpretations of a student's performance, they begin to consider many differing interpretations of student readings of text. Because the cases are constructed to engage common misconceptions and provoke authentic questions about student reading and thinking, during a case discussion, individual teachers will give voice to conflicting perceptions. Case discussions are carefully facilitated to help teachers see the value of classroom conversations about reading as formative assessment data and practice making evidence-based claims about students' reading and learning strengths and needs rather than basing instruction on preconceptions and assumptions about student abilities. Interpreting student thinking on the basis of observations during reading activity helps teachers develop insights into student

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

thinking during the case and later develop new insights into their own students as they listen to their students discuss science readings in their classrooms or examine their reading-related work.

In addition to these analytical tools, the professional development provides teachers with an experiential knowledge base of literacy routines, strategies, and tools that support teachers as they develop adaptive expertise that allows them to solve problems flexibly. This instructional tool kit, while situated in ongoing inquiry activities and explorations, is designed to meet teachers' needs for concrete and practical solutions to the everyday problem of students' limited comprehension of course texts. At the same time, teachers are invited to reflect on and critique instructional techniques and to adapt them to science teaching and their own students, grounded in principles of instruction in the Reading Apprenticeship instructional framework. Finally, to foster connections to reform initiatives and expectations operating in their school sites, teachers are given opportunity and support to explore the fit between the literacy approaches they already utilize, those they are learning in the professional development sessions, and the curriculum standards for which they are accountable.

Potential Contributions of this Study

Previous studies employing the Reading Apprenticeship professional development model have shown it to be effective in changing teachers' knowledge and classroom practice and increasing students' literacy achievement (Greenleaf & Schoenbach, 2001, 2004). These studies of the model, using a mix of qualitative and quasiexperimental methodologies, have suggested that participating teachers change their beliefs about the role of reading in content-area instruction, enlarge their conceptions of literacy (enriching what are often impoverished views of the complexities involved in reading and comprehending texts), expand their repertoire of pedagogical practices to support reading development, implement new instructional strategies, increasingly view subject-area reading tasks from the point of view of learners, and listen to students with new insights into their process of learning (Greenleaf & Katz, 2004; Greenleaf & Schoenbach, 2001, 2004). The current study extends previous research by explicitly examining links between professional development embedding core features of effective teacher development, teacher development of adaptive expertise needed to meet the high standards envisioned in such documents as the Common Core Standards, and student achievement in reading and science. Furthermore, the study tests these links utilizing rigorous research methodologies, in schools serving high-need students, using distal measures that acknowledge the "high stakes" at play in the current educational research and reform environment.

Professional development to implement Reading Apprenticeship in biology served as the intervention for this study. On the basis of previous studies, we hypothesized that through this professional development experience, teachers would develop new knowledge and resources about text, science

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

reading, and student thinking that would support their teaching of reading in science. With practice, teachers would learn to deploy these resources more flexibly, on demand, as students need them. In turn, students would practice these thinking tools as authentic and relevant responses to real reading situations and to make sense of science text as they build knowledge of the topic rather than as a set of fixed exercises in isolation from sense-making or knowledge building. Through literacy practice situated in authentic learning, we expected both teachers and students to stretch beyond their current ability and gain more expertise and capacity.

Research Hypotheses

We designed a randomized controlled study to test these hypotheses, examining impacts of the professional development model on teachers' instructional practices as well as on student literacy and science learning.

Hypothesis 1: Teachers participating in the Reading Apprenticeship professional development program will exhibit greater increases in knowledge and skills regarding the integration of literacy and science and will demonstrate greater integration of literacy into their instructional practice than teachers in control classrooms.

Hypothesis 2: Students in experimental classrooms will demonstrate greater increases in science understanding, reading proficiency, and engagement in science learning than their counterparts in control classrooms.

Study Methods

To test these hypotheses, a true, group-randomized, experimental design (Cook & Campbell 1979; Murray 1998) was conducted to control for most threats to internal validity in order to assess the impact of Reading Apprenticeship professional development on high school biology teaching and learning. The study targeted schools serving traditionally low-achieving students from groups that are underrepresented in the sciences. A multirole study team was assembled to include developers of the intervention and their research and professional development staff, science content experts, and independent evaluators. Roles were carefully delineated such that developers and their staff were involved in designing the professional development and instrumentation for the study, while independent evaluators were responsible for instrument scoring and analysis.[1] Data firewalls prevented developers from having access to data files.

Recruitment and Random Assignment of Schools and Teachers

The target population was high school biology teachers and their students in public high schools across California serving populations of students historically underrepresented in the advanced sciences. Schools, not

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

teachers, served as the unit of randomization to minimize contamination of the control group through teacher interaction. Prior to randomization, teachers volunteering for the study were recruited from each participating high school. To control for the effect of experience, all teachers qualifying for the student had to be credentialed in biology and had to have taught for at least 2 years at the initiation of the data collection phase. Thus, all teacher participants in the study were volunteers and presumably interested in the potential of literacy and science integration. Once teachers and schools were recruited, participating high schools were matched with similar schools on the basis of the California Department of Education's 2004 School Characteristics Index (SCI), a composite index representing a school's demographic composition (California Department of Education, 2000).

The SCI is based on the following factors: student mobility (percentage of students who first attended school in current academic year), ethnicity (percentage of students in seven ethnic-race categories), average parental education, percentage receiving subsidized meals, percentage of teachers fully credentialed, percentage of teachers with emergency credentials, percentage of English language learners, average class size, and year-round school status. In creating the index, each factor was weighted proportional to its relationship to the California's Academic Performance Index, based on a linear regression model.

Schools and the participating teachers within them were then randomly assigned to one of two different groups: an experimental group and a waitlisted control group (see Table 1). Random assignment was conducted in six waves in the spring of 2005 to allow more intervention teachers enough time to adequately incorporate their summer professional development training into their summer schedules. In total, 105 biology teachers in 83 schools were recruited, with 56 teachers (43 schools) assigned to the treatment group and 49 teachers (40 schools) assigned to the control group.[2] Participating teachers exhibited approximately 8.5 years of experience teaching science and had been assigned to the school for 6.25 years, on average. Approximately 62% had a postgraduate degree. Moreover, during the final intervention implementation year, approximately 84% of control teachers reported that they participated in some kind of training related to science or science education, 44% reported that they participated in summer institutes related to science education, and 36% attended college courses related to science education (compared to 26% among intervention teachers). Overall, participating teachers in both the intervention and control groups exhibited high levels of experience and engagement in professional development activities targeting science instruction.

Professional Development Intervention

Experimental condition. Teachers in schools randomly selected to be in the experimental condition received a total of 10 days of professional development in Reading Apprenticeship and support to integrate science content

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 1
Experimental Design

	Year 1, 2005–2006			Year 2, 2006–2007		
Group	Summer	Fall	Spring	Summer	Fall	Spring
Teachers						
9th/10th biology						
Group 1	O	TxU	O		TxU	O
Group 2	O PD	RA	O	PD	RA	O
Students						
8th grade						
Group 1			O			
Group 2			O			
9th grade						
Group 1			O		TxU	O
Group 2			O		RA	O
10th grade						
Group 1					TxU	O
Group 2					RA	O

Note. O = observations or measurement points; PD = Reading Apprenticeship professional development; RA = classroom implementation of Reading Apprenticeship; TxU = treatment-as-usual condition.

and reading instruction. The 10 days of professional development were led by certified Reading Apprenticeship professional development providers and utilized Reading Apprenticeship inquiry tools and approaches to professional development. In addition to these inquiry tools, the institute was designed to integrate literacy and topics in the high school biology curriculum, engaging teachers in experiential learning with the inquiry approaches and pedagogies they were being asked to implement in their science teaching but at a level of complexity suitable for adult and experienced science learners (Loucks-Horsley et al., 2003; Loucks-Horsley & Matsumoto, 1999). As described earlier, the professional development curriculum involved teachers experientially in using metacognitive routines, such as think-aloud (Kucan & Beck, 1997), and metacognitive logs for reading and for science investigations (Schoenbach et al., 1999); teacher modeling of reading and reasoning processes with think-aloud and text annotation (Greenleaf, 2006); methods of orchestrating and conducting collaborative small-group work, such as think-pair-share, jigsaws, and other group protocols involving comprehension routines, such as ReQuest (Manzo, 1969) and Reciprocal Teaching (Palinscar & Brown, 1984); and engaging in extended reading opportunities with varied sets of texts on particular topics. These instructional approaches

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

were tightly integrated with core units of study in biology to illustrate science and literacy integration. Metacognitive reflection on the impact of these learning opportunities and specific pedagogies, for their own and their students' learning, followed the integrated lessons. The 10 days of professional development were spread over a year, beginning in summer of Year 1, with follow-up midyear, and a final session before the start of Year 2. Figure 1 summarizes the content and structure of the professional development. The Appendix (see the online version of the journal) gives a more detailed set of agendas for the 10 days of professional development.

In the summer of 2005, teachers participated in a 5-day professional development institute. Implementation of reading instruction in their biology classes began in the fall of 2005. The professional development coaches made use of informal interviews and/or e-mail interactions with these teachers to plan two follow-up days of training given during Year 1 (2005–2006 school year), targeting the teachers' emerging needs for support. A final 3-day professional development follow-up occurred in the summer of 2006, prior to the data collection year. Throughout the study, exchanges took place through a listserv moderated by the professional development coaches. To support implementation and to assure equal access across experimental sites to opportunities to read in science, teachers were provided funds and a list of reading materials to supplement their locally adopted textbooks. These materials constituted a classroom library of science magazines, trade books, fiction, and nonfiction selections linked to the biology topics and state curriculum frameworks.

Control condition. Teachers in schools randomly selected to be in the control condition were offered the Reading Apprenticeship professional development in the summer of Year 3, after classroom data collection was completed. During the first 2 years of the study, they implemented their usual teaching practices. Thus, the control group represents a treatment-as-usual condition, representing what students would normally receive at schools participating in the study. However, teachers in the control condition were also offered the library of supplemental reading materials given to intervention group teachers so that the difference between groups, if any, was not attributable to whether or not such materials were present in classrooms. The treatment-as-usual control group condition meant that we made no attempt to restrict or circumscribe typical district and school activities, such as curriculum or textbook adoptions, staff development, school reform or policies, and so on. Participation in staff development activities and other changes in conditions and circumstances in the schools were tracked and monitored in both intervention and control group sites as part of yearly survey administration.

Data Collection and Analysis Procedures

Study methods included multiple measures of both teachers' instructional practices that shape students' opportunity to learn in the classroom

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

5-Day Professional Development Institute Summer, 2005	Ongoing Support
Topic focus: cell biology, evolution, and genetics	5 Additional institute days 2 in February 2006 3 in August 2006
Reading activities and inquiries with biology text sets (sets of texts of varied reading complexity; varied types of texts, including graphs and diagrams)	Additional reading activities, science investigations, comprehension strategies, video inquiries
Science investigations (acids and bases, finch taxonomy, genetics lab)	Word-learning strategies
Metacognitive routines and conversation (think-aloud, text annotation, metacognitive logs, etc., followed by shared problem solving)	Curriculum embedded reading assessment, examining student work
Collaborative learning (Think-Pair-Share, Team Reads, ReQuest, Reciprocal Teaching, Jig-Saw)	Examining and linking literacy to high school biology standards for California
Comprehension strategies (using clarifying, questioning, summarizing, predicting, visualizing, word analysis, etc., for readings)	Collaborative lesson planning, sharing, and critique
Classroom video inquiries	Classroom libraries linked to biology curriculum
Student case studies/video inquiries	Listserv with professional community and ongoing moderation/mentoring from Reading Apprenticeship coach

Figure 1. **Professional development design: Collaborative inquiries over 10 days**

and student achievement, and they were designed to enable us to determine the extent to which these instructional methods might have different impacts for groups of students historically underrepresented in the sciences. Study measures included a set of pre- and postintervention survey assessments of teacher knowledge, beliefs, and instructional practices in science and literacy; postintervention interviews; samples of lesson assignments with accompanying student work in particular biology topics; student surveys; and pre- and postintervention assessments of student learning in biology and reading comprehension. Participating schools were spread across many districts in the large, geographically varied state of California. Sending researchers to these many sites to carry out multiple direct observations of all classrooms would have been prohibitively expensive in a study of this size. (Results from a small subset of treatment classrooms observed for anecdotal

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

purposes will be reported elsewhere.) Therefore, we designed data sources and data collection procedures to enable us to corroborate data from various sources, using robust proxy measures of implementation. For example, rather than relying simply on teacher self-reports in surveys and interviews, we also used a method of collecting and analyzing lesson assignments and student work samples that has been shown to serve as a good proxy for classroom observations (Clare & Aschbacher, 2001; Matsumura, 2003). Student surveys provided a check on teacher self-report regarding classroom practices as well as a measure of student engagement, self-efficacy in science reading, and motivation. Instruments were designed to measure similar constructs to facilitate corroboration across measures. We describe these instruments in more detail below.

Measures of Teacher Knowledge, Belief, and Instructional Practice

Teacher survey. Based on the theoretical constructs underlying the Reading Apprenticeship instructional framework and accompanying professional development model, the teacher survey was designed to assess six global constructs related to effective integration of literacy and biology instruction: (a) science reading opportunities, (b) collaboration, (c) metacognitive inquiry, (d) comprehension strategies instruction, (e) a feature of instruction called "negotiating success"—a focus on designing and modifying instruction and assessment supporting response to student learning needs—and (f) teacher beliefs about reading, learning and diversity (Table 2). The six constructs were further divided into 14 subconstructs reflecting aspects of the apprenticeship model: the range of science reading opportunities offered to students and the nature and degree of teacher modeling, guidance, and support for, as well as student opportunities for practice with, key reading and discourse routines, tools, strategies, and dispositions.

Related to each of these constructs, we developed a set of items describing instructional practices or, in the case of Construct 6, a set of value statements. Teachers responded on a 5-point Likert-type scale regarding the degree of emphasis they placed on the item or its frequency of use or the degree to which they agreed with the item. To pilot the survey, we administered it to a set of teachers not participating in the study and conducted a factor analysis. With few exceptions, items loaded on the constructs they were expected to, and nonloading items were omitted from the survey. The resulting survey constructs had reasonably good psychometric properties, as alpha levels show in Table 2. However, the negotiating success assessment and teaching philosophy constructs exhibited low levels of internal consistency reliability, with Cronbach's alpha levels ranging from .41 to .55.[3]

The teacher survey was administered to teachers before the professional development intervention, a year later at the end of Year 1 of the study, and again at the end of Year 2. The Year 2 survey serves as a posttest to the identical surveys taken in summer 2005 and summer 2006. Presurvey

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 2

Teacher Survey Constructs, Subconstruct Definitions, Example Items, and Cronbach's α

Construct	Subconstruct Definition	Example	α
Construct 1: Science reading			
1a. Texts (13 items)	The range of science reading materials used in instruction	How often do you use the following reading materials and texts in your instruction? (list of types follows)	.74
1b. Learning structures (6 items)	How reading assignments are carried out, with whom, and in what context	How often do your students read assigned materials silently in class?	.73
1c. Content (5 items)	The extent to which science content from reading materials is acquired through student meaning making (versus summarized by the teacher)	How often do you present the important text information verbally to make sure everyone gets it after reading?	.80
Construct 2: Collaborative activities			
2a. Teacher modeling (6 items)	The extent to which teachers modeled and supported collaboration instructionally	To what extent do you establish conversational routines to promote student to student talk?	.84
2b. Student practice (6 items)	The extent to which students had access to one another for core work	How often do your students critique and challenge one another's ideas or work?	.80
Construct 3: Metacognitive inquiry			
3a. Teacher modeling (5 items)	The extent to which teachers modeled metacognitive inquiry and reading routines	How often do you think aloud to model your own confusions and efforts to make sense of readings?	.77
3b. Student practice (7 items)	The extent to which students had opportunities to practice metacognitive inquiry and reading routines	How often do your students discuss confusions and ways to make sense of reading materials?	.87
Construct 4: Comprehension strategies			
4a. Teacher modeling (3 items)	The extent to which teachers provided modeling and explicit instruction in comprehension-supporting strategies	How often do you model the use of various reading comprehension strategies?	.91

(continued)

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 2
(continued)

Construct	Subconstruct Definition	Example	α
4b. Student practice (22 items)	The extent to which students had opportunities to practice comprehension-supporting strategies	How often do you ask your students to visualize and/or construct visual representations?	.91
Construct 5: Negotiating success			
5a. Instruction (9 items)	The extent to which teachers modify instruction on the basis of student need to promote successful engagement and learning	To what extent do you negotiate assignments, deadlines, and policies with students?	.74
5b. Assessment (7 items)	The extent to which formative assessment informs instruction	To what extent do you provide verbal feedback to students through formal and informal conferencing?	.50
Construct 6: Teaching philosophy			
6a. Reading (12 items)	The extent to which teachers view reading as valuable in biology learning	Spending class time reading runs counter to the goal of building science knowledge through hands-on science inquiry.	.47
6b. Learning (12 items)	The extent to which teachers believe student learning is a process of constructing new knowledge in relation to their prior conceptions, which is facilitated by interaction with others	Students can learn as much from one another as they can from the teacher.	.55
6c. Diversity (11 items)	The extent to which teachers believe learning differences and varied language and cultural backgrounds can be an asset in the classroom	Students should be able to speak English well before taking biology.	.41

responses were analyzed to determine whether there were initial differences between mean responses of treatment and control group teachers on each construct and subconstruct. To examine pre- and postintervention differences

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

between the treatment and control groups, we conducted a regression analysis using individuals' pretest responses as a covariate.

Teacher assignments. The use of teacher classroom assignments as an indicator of practice is a methodology developed by Center for Research on Evaluation, Standards, and Student Testing (CRESST) researchers (Aschbacher, 1999; Clare, 2000). Teachers are asked to submit sample lesson materials as well as student work from a particular assignment or unit of study. Along with the lesson materials (texts, handouts, etc.), teachers complete an extensive open-ended questionnaire about the sequence of instruction leading up to the assignment; the kinds of skills and strategies students were asked to demonstrate; what learning activities students engaged in, and how, and with whom, in carrying out the assignment; what instructional support they received; expectations for student performance; and how students would be assessed. Together, the elicited information, lesson materials, and student work samples are given quality ratings based on a rubric. CRESST research supports the validity and reliability of using classroom assignment ratings as an indicator of classroom practice quality and proxy for classroom observation (Clare & Aschbacher, 2001; Matsumura, 2003). Using assignment ratings to assess practice has added benefits of reducing burdens on both teacher time and data collection resources in comparison to other methods.

The original CRESST Teacher Assignment instrument focused on science content alone. For this study, we modified the original content dimensions to reflect Reading Apprenticeship's focus on metacognitive inquiry and added new literacy dimensions that measure the quality of the teacher's goals for student literacy learning in the assignment (their purpose, clarity, and elaboration), the degree and nature of the literacy challenge offered to students in the assignment (complexity of texts, degree of challenge in associated tasks, and degree of metacognitive challenge), and the degree and nature of support for literacy challenge present in the assignment (support for engagement with science text, support for metacognitive challenge).

For each construct, a 4-point scale (1 = *poor*, 4 = *excellent*) was used to rate the quality for each assignment for separate dimensions. Figure 2 shows an example rubric scale for the literacy dimension, Support for Engagement with Science Texts. The resulting rubric was designed to allow scorers to gauge from lesson assignment materials, student work, and accompanying teacher descriptions of instruction the nature of literacy instruction, opportunities for engagement with challenging science texts, metacognitive inquiry into reading and thinking processes, and teacher support for the cognitive and metacognitive demands of the literacy task.

Teachers submitted two class assignments with six corresponding samples of student work, representing high, medium, and low quality. The two assignments came from two different topics in biology, genetics and cell biology. In addition to the lesson plan and student work samples, each

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

4

Students are *well supported* in their engagement with texts through **previous or current** teaching approaches, such as

- teaching of reading comprehension strategies
- structuring of the reading activity into an adequate number of explicit steps (e.g., analyzing text schema in small groups before independent reading)
- opportunities to read in class
- making material and social resources available to aid in comprehension (e.g., reading logs and discussions)
- opportunities to work collaboratively

3

Students are *adequately supported* in their engagement with texts.

- The teacher provides—or has previously provided—students with an **adequate** amount of information on the use of reading comprehension strategies and incorporates time into the assignment for reading and the use of reading comprehension strategies.
- The activity is fairly well structured (e.g., scaffolded as necessary).
- Social and material resources may be provided to aid in comprehension (e.g., graphic organizers, consumable texts, and small-group interactions).

2

Students are *somewhat supported* in their engagement with texts.

- The teacher may provide—or may have previously provided—students with **minimal** information on reading comprehension strategies and may incorporate some time into the assignment for reading and the use of reading comprehension strategies.
- The assignment may not be appropriately structured and resources may not be provided to aid in comprehension (e.g., reading logs or peer support).

1

Students are *generally not supported* in their engagement with texts.

- The teacher does not provide students with any information on reading comprehension strategies through previous or current teaching approaches or incorporate time into the assignment for reading and the use of reading comprehension strategies or both.
- The assignment may be disorganized.
- Resources may be inadequate or not provided to aid in comprehension.

Figure 2. **Rubric scale for teaching assignment literacy dimension: Support for Literacy Challenge, Support for Engagement with Science Texts**

teacher also submitted an in-depth cover sheet for each assignment, in which they described aspects of the lesson, including any in-class support that was provided, reflection about the lesson implementation and success, and student engagement with the material.

A scoring team composed of researchers and teachers who had not participated in the study was trained to score, using anchor (example) lesson assignments and the rubric. Interrater reliability was established by scoring

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

anchor assignments in common and working toward consensus on all rubric scales. Once the training process was complete, each Teacher Assignment was scored by at least two raters on the rubric dimensions. Through discussion and using initial independent scores as a focus for these discussions, the raters established final consensus scores for all dimensions. Assignments with more than a 1-point difference on at least one dimension were scored by a third rater. The final assignment ratings represent the consensus score across the raters. In addition to these ratings, we also conducted a descriptive analysis of assignment content and activity to allow us to construct concrete pictures of the instruction in treatment and control classes.

Teacher interviews. Interviews were conducted using semistructured protocols aligned with the teacher survey content and constructs. The interviews focused on eliciting and probing the nature and degree of teachers' implementation of classroom practices targeted by the intervention. All intervention and control teachers were interviewed in the spring of the data collection year. The interviews took place by phone and were recorded for analysis. Interviews were subsequently rated on a 4-point rubric on five dimensions: reading opportunities, support for student reading engagement, metacognitive inquiry, reading comprehension routines, and collaboration. An exacting training process was developed to establish interrater reliability by scoring 10 randomly selected interviews in common and working toward consensus on all constructs. A sixth construct, inquiry science, was developed post hoc to the actual interviews and scored as a dichotomous outcome. Participants were given a score of 1 if there was significant evidence of regular inquiry science teaching practices within the interview and a score of 0 otherwise. The interview scoring rubric is quite lengthy and will be provided to readers on request. Figure 3 shows the score points for one rubric dimension, metacognitive inquiry, which is at the core of the intervention.

Measures of Student Learning Opportunities And Outcomes

Parental consent forms and student outcome measures, including Opportunity to Learn (OTL) surveys and state standardized test scores for the baseline and intervention year, were collected in one focus class for each participating teacher in the control and intervention conditions, to enable us to link baseline scores, intervention year scores, OTL surveys, and teacher implementation measures. Teachers were instructed to administer student surveys and assessments to third period, if they taught biology at that time, or to the period closest to it that they taught biology. Our intent was to maximize complete data and minimize absenteeism by avoiding first period or periods after lunch. Nonidentifiable student standardized test score data were collected for each participating teacher to broaden the sample and its representativeness.

Student OTL survey. On the basis of prior surveys developed at WestEd for the Performance Assessments in Science (PASS) assessments, student

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

METACOGNITIVE INQUIRY

The purpose of this construct is to measure opportunities for metacognitive inquiry into science reading and thinking processes, including teacher modeling, guidance, and support; student practice of metacognitive conversation and routines; teacher approach to challenges; and accountability and assessment of student metacognition.

4

Metacognitive Conversation: Reading ***frequently*** involves noticing, sharing, and problem-solving confusions, reading and thinking processes, and sense-making. Metacognitive conversation about reading and thinking processes ***takes center stage*** in an ***ongoing*** metacognitive conversation.

Teacher Instruction and Modeling of Metacognitive Processes, Routines, Tools, and Strategies: Teacher ***often teaches and models*** reading and thinking processes, routines, and strategies that support students to become self-monitoring and self-governing readers of science. May be a shift toward reduced modeling and instruction as the year progresses, but ***modeling continues*** when new challenges (e.g., texts and tasks) are introduced. Modeling and demonstration generally ***emphasize*** reading and thinking processes rather than procedures.

Student Practice: Students ***often practice*** metacognitive reading routines, tools, and strategies (most lessons).

Approach to Challenges: Teacher ***usually encourages and supports*** students to grapple with challenging texts, tasks, and concepts.

Accountability and Assessment of Metacognition: Teacher ***frequently collects and/or assesses*** students' reading and thinking processes on an ongoing basis, using two or more measures (e.g., reading logs, annotations, student talk). Focus of assessment may shift across year, for example, from focus on comprehension monitoring and strategies use to evidence of scientific reading and thinking, but teacher ***continues to assess*** reading and thinking processes as well as content.

3

Metacognitive Conversation: Reading ***sometimes*** involves noticing, sharing, and problem-solving confusions, reading and thinking processes, and sense-making (one to two times a week). ***May be more frequent but limited***, for example, confusions and understanding may focus primarily on right answers rather than on reading and thinking processes. Teacher ***sometimes engages students*** in metacognitive conversation about reading and thinking processes. Metacognitive conversation, ***though ongoing, may be limited*** to particular times or activities (e.g., "Metacognitive Mondays") rather than pervasive.

Teacher Instruction and Modeling of Metacognitive Processes, Routines, Tools, and Strategies: Teacher ***sometimes teaches and models*** reading and thinking processes, routines, and strategies that support students to become self-monitoring and self-governing readers of science. May be a shift toward reduced modeling and instruction but modeling and instruction are ***ongoing to some extent*** (once a week or so). Modeling and demonstration ***sometimes emphasize*** reading and thinking processes and sometimes emphasize procedures and correctness.

Student Practice: Students ***sometimes practice*** metacognitive reading routines, tools, and strategies (once or twice a week); or student practice, although more frequent, focuses on answers rather than on reading and thinking processes—for example, double-entry reading logs focused on facts and examples.

(continued)

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Approach to Challenges: Teacher ***sometimes encourages and supports*** students to grapple with challenging texts, tasks, and concepts; or grappling, although frequent, may involve ***limited teacher support*** in how to approach challenging materials.

Accountability and Assessment of Metacognition: Teacher ***sometimes collects and/or assesses*** students' reading and thinking processes, for example, after teaching a new routine or strategy. Focus of assessment may shift across year from focus on reading processes to content or correctness.

2

Metacognitive Conversation: Reading ***occasionally*** involves noticing, sharing, and problem-solving confusions, reading and thinking processes, and sense-making, for example, ***sporadically or at the beginning of the year***. Teacher ***occasionally engages students*** in metacognitive conversation about reading and thinking processes, for example, at the beginning of the year only, of narrow focus, through a single routine or tool, or used sporadically.

Teacher Instruction and Modeling of Metacognitive Processes, Routines, Tools, and Strategies: Teacher ***occasionally teaches and models*** reading and thinking processes, routines, and strategies, for example, ***sporadically or at the beginning*** of the year. Modeling and demonstration may occasionally focus on reading and thinking processes but ***generally emphasize procedures and correctness***.

Student Practice: Students ***occasionally practice*** metacognitive reading routines, tools, strategies, and conversations (a few times a month).

Approach to Challenges: Teacher ***occasionally*** has students grapple with challenging texts, tasks, and concepts, albeit with ***little teacher support***.

Accountability and Assessment of Metacognition: Teacher ***occasionally collects and/or assesses*** students' reading and thinking processes. Assessment may be ***sporadic or short-term***.

1

Metacognitive Conversation: ***Little or no opportunity*** for students to notice, share, or problem-solve confusions, reading and thinking processes, and sense-making. ***Little or no metacognitive conversation*** about reading and thinking processes. Teacher may view ***sharing confusions as "complaining" or discouraging*** to students.

Teacher Instruction and Modeling of Metacognitive Processes, Routines, Tools, and Strategies: Teacher ***does not teach or model*** reading and thinking processes, routines, or strategies. Teacher modeling and demonstration, if any, ***almost always focuses on procedures and correctness*** rather than on reading and thinking processes.

Student Practice: Students ***rarely or never practice*** metacognitive reading routines, tools, strategies, and conversations (once a month or less).

Approach to Challenges: Teacher ***rarely or never has students grapple*** with challenging texts, tasks, and concepts.

Accountability and Assessment of Metacognition: Teacher ***rarely or never collects and/or assesses*** students' reading and thinking processes.

Figure 3. **Rubric scale for teacher interview: Metacognitive inquiry dimension.**

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

reading surveys developed by Greenleaf and colleagues (Greenleaf et al., 2001), and CCSSO's Survey of Enacted Curriculum (www.ccsso.org/projects/Surveys_of_Enacted_Curriculum), we developed an OTL survey. The survey asked students about classroom practices related to the integration of literacy and biology, but it also included items related to student engagement, motivation, and students' perceptions of themselves as readers and learners. Six key constructs were assessed by the survey and were used as outcomes: (a) class emphasis on reading in biology, (b) integration of biology and literacy activity, (c) identifying as a reader, (d) student identity, (e) motivation in class, and (f) course consequences on reading science. Using pilot data, we used a series of exploratory and confirmatory factor analyses to validate the factor structure of the student survey items. Table 3 provides a map of individual survey items to each of these constructs and Cronbach's alpha levels for each construct. The survey was administered to students in both treatment and control groups in spring of 2006–2007, the intervention year.

State standardized test scores. To broadly assess student performance in biology and reading comprehension, we relied on available state-mandated criterion-referenced tests. Although we explored the use of additional standardized tests that would have been more sensitive to our intervention, such as the PASS test in biology or the Degrees of Reading Power test of reading comprehension, we discovered that schools and districts would not agree to administer a nonmandated test to students since so much instructional time was already devoted to testing in the state. We therefore used the California Standards Test (CST) Biology test to assess biology understanding and the CST English Language Arts (ELA) and Reading Comprehension subtest to measure reading comprehension.

The California state standardized tests in English and biology are not particularly well suited to the intervention of this study. The vast majority of biology items, for instance, are concept identification or factual recall questions on content that require very little reading. Conversely, while the English test requires reading, the vast majority of items focus on literature. Nevertheless, these tests, however distal a measure of student achievement in the areas targeted, represent both a readily available and critical measure of the impact of Reading Apprenticeship professional development, given the increasingly high stakes attached to state standardized measures.

Standardized test data collection proved quite difficult across multiple school districts with varied research capabilities and was not completed until the fall of 2009. Data were requested from every district from which we had had a teacher participate, even if the specific teacher was not retained in the study. Our final data collection resulted in two types of data. For students from whom we had obtained parental consent, we collected linked, longitudinal test score data. We also collected anonymous cross-sectional data linked to teachers, but not to specific students, for those from whom we did not obtain parental consent. For linked students we collected CST ELA and ELA Reading Comprehension for 2005–2006 prior to student exposure to study

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 3

Student Opportunity-to-Learn Survey Constructs, Items, and Cronbach's α

Construct and Item	α
(1) Class emphasis on reading in biology	.80
1. Reading a wide variety of science materials (textbooks, lab procedures, etc.)	
3. Working together to figure out the meaning of the readings	
4. Listening and responding to one another's ideas	
5. Learning to read, write, listen, and talk about science	
6. Taught ways to make science reading interesting and motivating for students	
7. Taught different strategies to help students understand science reading better	
8. Taught students how to read charts, graphs, tables, and illustrations	
9. Shared what is going on in the teacher's mind while the teacher reads science material	
11. Encouraged students to borrow one another's ideas	
(2) Frequency of student integration of biology and literacy activity	.77
14. Spent class time reading	
15. Worked with partners or groups on reading assignments in class	
16. Practiced reading comprehension strategies with science materials	
17. Shared difficulties and ways you solved reading comprehension problems	
18. Figured out vocabulary in science reading materials	
19. Analyzed the way science materials are written and organized	
(3) Perceived course consequences on identifying as a reader	.74
29. Understanding yourself better as a reader and learner	
30. Making you curious to read about other things in science	
31. Seeing yourself as a reader.	
(4) Perceived course consequences on student identity	.80
32. Being a more serious student	
33. Thinking about your future educational goals	
34. Making you interested in taking more science classes	
38. Thinking of yourself as a capable student	
39. Feeling like you can succeed in more challenging classes	
40. Seeing your education as important	
(5) Motivation in class	.84
21. Completed reading assignments	
22. Enjoyed completing a reading assignment . . . that required a lot of thinking . . .	
23. Put forth a great deal of effort when doing your biology reading	
25. Tried to really understand biology reading assignments in this class	
26. Felt motivated to work harder than usual on reading assignments in this class	

(continued)

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 3
(continued)

Construct and Item	α
27. Wanted to do a good job on reading assignments	
28. Became really interested in the science reading assigned in this class	
37. Being willing to tackle challenging reading materials	
(6) Perceived course consequences on reading science	.90
35. Understanding science materials better when you read	
36. Given you more confidence that that you can read and do science	
41. Learning science better	
42. Understanding science concepts better	
43. Feeling like you can be more successful reading in other science classes	
44. Feeling more positive about reading science	
45. Having a more positive attitude about reading in general	

teachers and for 2006–2007. The 2005–2006 CST measures were used as baseline covariates in the longitudinal impact analysis models. For students who are unlinked in the data set, we collected CST ELA, ELA Reading Comprehension, and Biology scores for students in participating teachers classrooms in 2004–2005, 2005–2006, and 2006–2007. The 2004–2005 anonymous student test score data were aggregated to the teacher level and were used as covariates in our cross-sectional student test score impact analysis models.

The ELA and Reading Comprehension tests are not vertically scaled and thus do not have the same meaning across different grade levels. To convert the scores to an identical metric so that test score data from all of the grades can be analyzed together, within each grade, test score data were normalized by subtracting the sample mean from each student's score and dividing by the sample standard deviation. Normalized in this way, the test score data represent the *relative ranking* of students in the analytic sample rather than the absolute level of performance, and the impact estimates reflect the standardized effect estimate.

Table 4 shows the schedule for the various data collected to gauge evidence of teacher implementation and student learning outcomes.

Results

Retention of Schools and Teachers

Figure 4 shows the number of teachers and schools randomly assigned to treatment and control groups as well as the data retention rates for each

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 4
Data Collection Schedule

	Baseline	Year 1	Year 2
	2004–2005	2005–2006	2006–2007
Teacher practice			
Teacher surveys (instructional beliefs/ practice)	Summer	Summer	Summer
Teacher assignments (genetics, cell biology)			Fall/Spring
Teacher interviews			Spring
Student Opportunity to Learn and outcome measures			
Student Opportunity to Learn surveys			Spring
State test scores (biology and English language arts), longitudinal data		Spring	Spring
State test scores (biology and English language arts), cross-sectional data	Spring		Spring

data source. As shown in Figure 4, 89% of treatment teachers and 76% of control teachers provided responses on the baseline teacher survey, 79% and 76% of treatment and control teachers participated in the 1st-year post-implementation teacher survey, and 59% and 53% participated in the 2nd-year post-implementation survey. Return rates for other types of data after the 2nd study year were similar to those for the 2nd-year post-implementation survey. Student longitudinal data and student OTL survey data were secured from approximately 50% of randomly assigned teachers, teacher interviews were conducted with 55% of teachers, and lesson assignment data were collected from 63% of treatment teachers and 45% of control teachers. Cross-sectional student test score data were collected from 64% of treatment teachers and 51% of control teachers. The school participation chart at the bottom of Figure 4 shows similar data return rates as that for teachers.

Equivalence of Treatment and Control Groups

Although data attrition levels were fairly high, attrition patterns were similar for treatment and control schools. Exceptions to this were apparent for the student cross-sectional data, the student OTL surveys, and the lesson assignment data, with higher data return rates exhibited for treatment teachers than for control teachers. To describe treatment and control group equivalence (or lack thereof) at the time of random assignment and at subsequent data collection periods, we examined school, teacher, and student characteristics by data source. Table 5 shows school characteristics by treatment and control status for the randomized sample, school and teacher characteristics

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

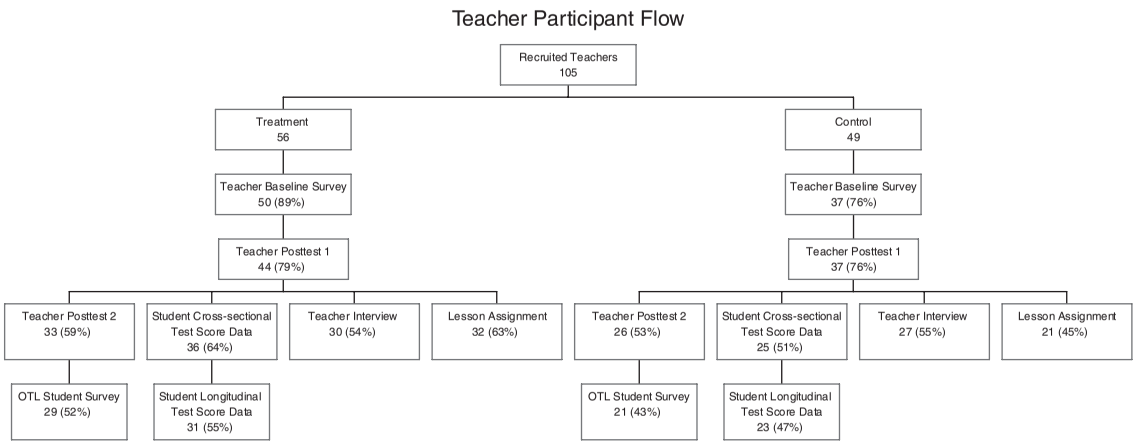


Figure 4. **Teacher and school retention by data source.**

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

for the initial (preinstitute) teacher survey sample, and school and teacher characteristics for the student OTL survey sample. Overall, the randomized and initial teacher survey samples show a high degree of similarity, with few meaningful differences in school performance and demographic characteristics. The student OTL sample, which is composed of about 50% of randomized schools and teachers, exhibits more evidence of treatment and control group nonequivalence than the teacher pretest sample, but none of the differences is statistically significant. Treatment schools had about 30% more English learners than control schools (21% vs. 16%), and participating teachers in treatment schools averaged about 1.8 more years of science teaching experience than their control group counterparts (9.3 vs. 7.5 years).

Table 6 shows preintervention characteristics of students in treatment and control schools by longitudinal test score, cross-sectional test score, and student OTL survey samples. Note that parental consent was required to collect student-level longitudinal test score and OTL data, so group differences in student characteristics reflected in the first and third panel of Table 6 could be due to differences in teacher participation rates, student participation rates, or both factors. Group differences in preintervention characteristics in the cross-sectional panel are most likely due to differences in teacher participation rates only.

No statistically significant differences between treatment and control schools were present, but as indicated by the longitudinal test score sample, treatment schools had higher proportions of English learners (42% vs. 25%) and Latinos (53% vs. 29%) and lower proportions of White students (16% vs. 33%). Treatment schools also exhibited baseline test scores that were between one fifth and one fourth of a standard deviation lower than those in control schools. This provides some evidence that participation of Latino students, English learner students, and students with lower standardized test scores was less likely in control schools than in treatment schools, but these differences could have arisen by chance factors alone. Also note that the proportion of English learner participants as reflected by the student-level characteristics in Table 6 was higher than that reflected by the school aggregate characteristics in Table 5 (42% vs. 21% in treatment schools), suggesting that the classrooms in schools that participated had greater proportions of English learners than the student body as a whole.

The treatment and control differences evident in the longitudinal test score sample are less pronounced in the cross-sectional sample, although treatment schools in the cross-sectional sample still had higher proportions of English learners (43% vs. 35%) and Latinos (44% vs. 35%) than control schools. Because the anonymous cross-sectional data were linked to teachers, but not to specific students, we did not have preintervention test score data for the students in the cross-sectional sample. Instead, we have anonymous preintervention data for 2004–2005 students served by the teachers in the analytic sample. As shown in the middle panel in Table 6, prior to the intervention, the students served by treatment teachers exhibited

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 5

School and Teacher Characteristics by Treatment and Control Status for Randomized Sample, Initial Teacher Survey Sample, and Opportunity to Learn (OTL) Survey Sample

Characteristic	Treatment	Control	Difference	*p* Value	Difference/*SD*
Randomized sample					
School characteristics (83 schools)					
API	653.39	649.46	3.93	.82	0.05
API state rank	4.93	4.89	0.03	.96	0.01
Science CST (NCE)	46.79	46.67	0.12	.94	0.02
Free/reduced-price meals	37.20	37.97	–0.78	.87	–0.04
African American (%)	11.27	9.24	2.02	.35	0.21
Hispanic (%)	39.59	44.05	–4.46	.42	–0.18
Asian (%)	10.13	8.90	1.23	.68	0.09
White non-Hispanic (%)	31.62	32.13	–0.51	.93	–0.02
English learners (%)	19.02	17.29	1.73	.57	0.13
Initial teacher survey sample					
School characteristics (70 schools)					
API	648.21	655.00	–6.79	.71	–0.09
API state rank	4.72	5.17	–0.44	.52	–0.16
Science CST (NCE)	46.20	47.19	–0.99	.54	–0.15
Free/reduced-price meals	38.03	35.80	2.23	.68	0.10
African American (%)	11.25	9.40	1.85	.44	0.19
Hispanic (%)	41.27	42.35	–1.07	.86	–0.04
Asian (%)	9.76	9.91	–0.15	.96	–0.01
White non-Hispanic (%)	30.02	31.87	–1.86	.76	–0.07
English learners (%)	19.44	16.97	2.48	.48	0.18
Teacher characteristics (87 Teachers)					
Female	0.62	0.75	–0.13	.20	–0.28
Biology major	0.26	0.32	–0.07	.59	–0.14
Years teaching science	9.39	8.32	1.07	.35	0.20
Years in school	6.58	6.14	0.44	.77	0.09
Student OTL survey sample					
School characteristics (42 schools)					
API	640.36	651.76	–11.39	.63	–0.15
API state rank	4.50	5.05	–0.55	.54	–0.19
Science CST (NCE)	45.54	47.25	–1.72	.41	–0.27
Free/reduced-price meals	38.70	35.89	2.81	.69	0.13
African American (%)	10.15	8.97	1.18	.68	0.13
Hispanic (%)	46.56	42.79	3.77	.64	0.15
Asian (%)	8.26	12.11	–3.85	.41	–0.42

(continued)

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 5
(continued)

Characteristic	Treatment	Control	Difference	*p* Value	Difference/*SD*
White non-Hispanic (%)	28.03	29.95	–1.92	.79	–0.08
English learners (%)	21.05	15.95	5.10	.23	0.36
Teacher characteristics (50 teachers)					
Female	0.55	0.71	–0.16	.25	–0.33
Biology major	0.25	0.42	–0.17	.24	–0.36
Years teaching science	9.28	7.48	1.80	.23	0.34
Years in school	6.45	5.38	1.07	.51	0.23

Note. The *p* values are based on multilevel regression models in which treatment group status is included as a covariate. Effect sizes are calculated by dividing group difference by the pooled standard deviation. API = Academic Performance Index; CST = California Standards Test; NCE = normal curve equivalence.

substantially lower biology and reading comprehension test scores than those served by control teachers, with group differences of between 40% and 50% of a standard deviation. Overall, for both sets of test score samples, the results suggest that differential attrition may have led to treatment and control group nonequivalence. However fewer treatment and control school differences were apparent in the student OTL survey sample. To account for potential biases brought about by differential attrition, the student standardized test score impact analyses models presented below include controls for aggregated (teacher-level) student biology and reading comprehension test scores assessed prior to the intervention (2004–2005).

Analysis of Treatment Effects

To estimate program impacts, outcomes for teachers and students in treatment classrooms were compared to the outcomes for their counterparts in control classrooms. We analyzed the effectiveness of intervention using hierarchical regression models to account for clustering of the data by school (Goldstein, 1987; Murray, 1998; Raudenbush & Bryk, 2002). In each of the impact analyses, we controlled for baseline (pretest) measures of outcome variables when available, randomization strata (i.e., pairs), and student-level covariates (race-ethnicity, gender, and language status) when analyzing student outcomes. In both the teacher and student outcome models, a random effect for school was included to account for the nesting of observations within schools, the unit of assignment. In addition to examining main effects for program impacts on student performance outcomes, we also examined differences in impacts for the following subgroups: (a) English learners and English-proficient students (test scores only), (b) females and males, (c) racial-ethnic groups, and (d) and low- and high-performing students. For the outcomes assessed with student OTL surveys, we examined differential

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 6

Student Characteristics by Treatment and Control Status for Longitudinal Test Score Sample, Cross-Sectional Test Score Sample, and Opportunity to Learn (OTL) Survey Sample

Student Characteristic	Treatment	Control	Difference	*p* Value	Difference/*SD*
Longitudinal test score sample (45 schools)					
Female	0.44	0.46	–0.02	0.34	–0.04
English learner	**0.42**	**0.25**	**0.17**	**0.36**	**0.35**
African American	0.08	0.08	0.00	0.70	–0.02
Asian	0.11	0.16	–0.05	0.89	–0.16
Latino	**0.53**	**0.29**	**0.24**	**0.34**	**0.50**
Other	0.13	0.14	–0.01	0.26	–0.03
White	**0.16**	**0.33**	**–0.17**	**0.07**	**–0.43**
English language arts CST 05 (std)	–0.10	0.16	–0.26	0.78	–0.26
Reading comprehension 2006 (std)	–0.08	0.11	–0.18	0.53	–0.19
Mathematics CST 2006 (std)	–0.10	0.17	–0.27	0.55	–0.27
Cross-sectional test score sample (51 schools)					
Female	0.46	0.47	–0.01	0.61	0.01
English Learner	0.43	0.35	0.08	0.12	0.16
African American	0.10	0.10	0.00	0.54	0.02
Latino	0.44	0.35	0.09	0.26	0.18
White	0.21	0.27	–0.05	0.52	–0.15
Teacher biology CST 2005	**321.72**	**329.56**	**–7.84**	**0.16**	**–0.52**
Teacher reading comprehension 2005 (std)	**–0.07**	**0.04**	**–0.11**	**0.08**	**–0.44**
Student OTL Survey Sample (42 Schools)					
African American	0.08	0.08	–0.01	0.73	–0.02
Asian	0.06	0.10	–0.04	0.48	–0.15
Latino	0.45	0.39	0.07	0.42	0.13
Other	0.20	0.22	–0.02	0.76	–0.05
Non-English speaker	0.36	0.27	0.09	0.15	0.20

Note. The *p* values are based on multilevel regression models in which treatment group status is included as a covariate. Bolded numbers correspond to treatment and control differences that may be substantively meaningful. Effect sizes are calculated by dividing group difference by the pooled standard deviation. The std in parenthesis means that the outcome has been standardized to have a mean of 0 and a standard deviation of 1. CST = California Standards Test.

impacts across student gender, race-ethnicity, and student home language (English and non-English). The study was designed to measure the impact of the intervention on teacher knowledge and classroom practice as well as student achievement. We report these impacts in two sections, below.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Teacher Outcomes: Integration of Literacy Into Instructional Practice

Teacher surveys. Analysis of pre- and postintervention surveys at the end of Year 2 offered evidence that the intervention had produced increased teacher facility in integrating biology and literacy teaching. These results are presented in Table 7. To summarize these results, there is little evidence that the means on preintervention survey measures differ by treatment and control status. Teachers in the intervention and control conditions reported similar instructional practices and teaching philosophies at the start of the study. However, we found significant differences favoring the treatment group relative to the control group on 8 of the 14 subconstructs, as shown in Table 7, at the end of the intervention year, namely, science reading, content; collaborative activities, teacher modeling; collaborative activities, student practice; metacognitive inquiry, teacher modeling; metacognitive inquiry, student practice; comprehension strategies, teacher modeling; teaching philosophy, reading; and teaching philosophy, learning.

Teacher responses to surveys at the end of the study thus show differences between intervention and control teachers in both their knowledge about the role reading plays in learning and in their repertoire of instructional practices. According to teacher reports on the survey, intervention classrooms are distinguished from control classrooms in the degree to which students—rather than teachers—are more frequently doing the work of comprehending, they receive greater teacher support for carrying out this work, and this support frequently takes the form of metacognitive inquiry into reading and thinking processes.

The strength of the differences between the intervention and comparison classrooms on these survey outcomes is quite strong, with effect sizes ranging from 0.48 to 1.14 standard deviation units, the majority in the range of 0.69 to 0.77. Moreover, intervention group teachers reported higher levels on two additional constructs: science reading opportunities, learning structures (p = .09)—indicating that students accessed the content of science texts through reading and class discussion rather than through teacher lecture—and negotiating success, assessment (p = .08)—indicating more use of assessment to drive instruction, although these increased levels were not statistically significant at conventional levels. Ancillary analyses indicated that the differences in postsurvey outcomes are unlikely due to treatment and control group differences in sample selectivity, as no differences were apparent between treatment and control group teachers on pretest measures among the sample of teachers with nonmissing 2nd-year postsurvey data.

Teacher assignments. Table 8 shows means on the 4-point scoring rubric for the cell biology and genetics classroom assignment ratings. For the cell biology assignment, teachers in the intervention group received higher ratings on cognitive challenge (p value = .05) and support for cognitive challenge (p value = .10) in the literacy dimension. These ratings were approximately one half of a standard deviation higher than the ratings of control

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 7

Teacher Survey Differences by Treatment and Control Group

Teacher Survey Year 1 to Year 2	Treatment	Control	Difference	*p* Value	Difference/*SD*
Student reading opportunities, texts	3.10	3.15	–0.05	.60	–0.12
Science reading opportunities, learning structure	3.09	2.85	0.24*	.09	0.39
Science reading opportunities, content	3.17	2.89	0.28**	.02	0.61
Collaboration, teacher modeling	3.37	2.77	0.60**	<.01	0.76
Collaboration, student practice	3.16	2.75	0.42**	<.01	0.69
Metacognitive inquiry, teacher modeling	3.18	2.63	0.55**	<.01	0.76
Metacognitive inquiry, student practice	3.14	2.24	0.90**	<.01	1.14
Comprehension strategies, teacher modeling	3.64	2.86	0.78**	<.01	0.77
Comprehension strategies, student practice	3.09	2.88	0.21	.12	0.35
Negotiating success, instruction	3.64	3.42	0.21	.12	0.40
Negotiating success, assessment	3.10	2.88	0.22*	.08	0.42
Teaching philosophy, reading	4.53	4.34	0.19**	.03	0.48
Teaching philosophy, learning	4.59	4.25	0.34**	<.01	0.76
Teaching philosophy, diversity	4.05	4.17	–0.12	.28	–0.24

Note. Data are based on regression-adjusted multilevel regression models that include pre-intervention measures of each outcome variable and dichotomous indicators for randomization strata as covariates. Effect sizes are calculated by dividing estimates by the pooled standard deviation of the outcome variable. Sample size consists of 59 teachers in 49 schools.
*Treatment-control difference is statistically significant ($p < .10$).
**Treatment-control difference is statistically significant ($p < .05$).

teachers. For the genetics assignment, teachers in the intervention group received higher ratings for support for cognitive challenge in both the content and literacy domains.

In addition to rubric scoring, we carried out an analysis of these assignments to describe the general lesson characteristics. We also analyzed assignments for

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 8
Teaching Assignment Differences by Treatment and Control Group

Teaching Assignment	Treatment	Control	Difference	*p* Value	Difference/*SD*
Cell biology					
Mean content scores					
Goals	2.95	3.23	–0.27	.25	–0.37
Cognitive challenge	2.69	2.74	–0.05	.84	–0.06
Support for cognitive challenge	2.80	3.09	–0.30	.19	–0.39
Mean literacy scores					
Goals	2.38	2.04	0.35	.13	0.43
Cognitive challenge	2.48	2.14	0.34**	.05	0.58
Support for cognitive challenge	2.50	2.10	0.40*	.10	0.46
Genetics					
Mean content scores					
Goals	3.07	2.77	0.30	.20	0.39
Cognitive challenge	2.93	2.69	0.24	.27	0.35
Support for cognitive challenge	3.20	2.76	0.45**	.02	0.69
Mean literacy scores					
Goals	2.65	2.21	0.44	.14	0.48
Cognitive challenge	2.43	2.18	0.25	.28	0.37
Support for cognitive challenge	2.94	2.31	0.63**	.02	0.79

Note. Data are based on regression-adjusted multilevel regression models that include dichotomous indicators for randomization strata as covariates. Effect sizes are calculated by dividing estimates by the pooled standard deviation of the outcome variable. Sample size consists of 51 teachers in 42 schools.
*Treatment-control difference is statistically significant ($p < .10$).
**Treatment-control difference is statistically significant ($p < .05$).

Reading Apprenticeship–specific characteristics, such as the use of metacognitive routines.

All teachers in the study regardless of treatment or control used some type of text as part of the assignment's information input phase. These texts included journal articles, lab write-ups, textbooks, and others. The types of texts used were similar between treatment and control assignments, with one notable exception. Control teachers used PowerPoint presentations or lecture notes as texts more frequently than treatment teachers, with 27% versus 7% of submitted assignments, respectively. Similarly, control teachers gave more lectures as part of submitted assignments (26% vs. 20% of the assignments of treatment teachers) and more "explicit instruction"—often a gloss for telling students information they need to know (71% of control lessons vs. 34% of treatment lessons).

In contrast, treatment assignments provided higher frequencies of reading engagement support than control assignments (80% vs. 48% of submitted lessons), were more likely to include collaborative work (44% small groups and

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

48% pair groups compared to 27% and 29% for control lessons), and included more student meaning-making activities, such as discussion (65% of assignments compared to 40% for controls), note taking (61% of assignments vs. 40% for controls), and use of reading comprehension strategies (80% of assignments vs. 55% for controls). A notable difference between treatment and control assignments was the frequent use of a written version of think-aloud in which students annotate as they read to reflect and support their reading and thinking processes. Across cell biology and genetics, nearly 50% of the treatment assignments utilized this strategy versus 16% of the control assignments. Thus, analysis of general lesson characteristics and Reading Apprenticeship–specific strategies in the teacher assignments supported differences in classroom practices reported by teachers themselves in surveys and interviews.

Teacher interviews. Table 9 shows mean ratings for treatment and control teachers. The results indicate that teachers in the intervention group exhibit substantially higher interview ratings than their counterparts in the control group in the areas of teacher support for student reading engagement, metacognitive inquiry in reading and thinking processes, reading comprehension routines, and collaboration. Although the kinds of reading students are asked to do—the range of texts and volume of reading expected—did not differ in intervention and control classrooms at conventional levels of statistical significance ($p = .12$), the degree and type of support for student engagement with those texts, the explicit teaching and modeling and guided practice using specific comprehension routines, and the amount of collaboration around reading in these classrooms do differ significantly. Moreover, these differences are quite large in magnitude, ranging from 0.90 to 1.15 standard deviation units.

The effects of these differences for students are illustrated in the following interview responses. In a typical response to the interview question, "How often, if at all, do students in this class discuss problems they have with science reading materials and work together to clarify what they don't understand?" a treatment teacher who received a score of 3 on the 4-point rubric on the subconstruct metacognitive conversation replied,

> Not always formal, but I say, "Can you tell them how you got that, or where or why?" That's thinking processes. For breaking things down, relating to prior knowledge, teaching them that learning is building on prior knowledge, I've been up front with them about that.

In contrast, a control teacher who received a score of 1 replied,

> I have no clue because it's not a formal part of the class; they would be doing it informally. How often do they talk to me about it? Every chapter I hear students saying they didn't understand this and they read it and can't find the answers to something, some questions I've asked, so it's a lot.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 9
Teacher Interview Differences by Treatment and Control Group

Construct	Treatment	Control	Difference	*p* Value	Difference/*SD*
Reading opportunities	2.85	2.57	0.28	.12	0.46
Support for student reading engagement	2.97	2.19	0.78**	<.01	0.90
Metacognitive inquiry	2.54	1.65	0.89**	<.01	1.15
Reading comprehension routines	2.92	2.06	0.86**	<.01	0.95
Collaboration	2.98	2.22	0.76**	<.01	1.11
Inquiry	0.17	0.13	0.04	.46	0.13

Note. Data are based on regression-adjusted multilevel regression models that include dichotomous indicators for randomization strata as covariates. Effect sizes are calculated by dividing estimates by the pooled standard deviation of the outcome variable. Sample size consists of 57 teachers in 48 schools.
*Treatment-control difference is statistically significant ($p < .10$).
**Treatment-control difference is statistically significant ($p < .05$).

Likewise, a treatment teacher who received a 3 on the subconstruct student practice of metacognitive routines, tools, and strategies reported,

> There is more student participation, students generating information they are supposed to get. Sometimes I have charts they need to fill out while they are reading. The chart has them list something they want to talk about in class because they don't understand it, they've made a connection, there's various reasons why. They will copy a phrase onto chart, say why they want to discuss. Another column where they identify what type of strategy they could use to get through on their own. After they do it individually we discuss as a class. It takes the place of what would have been a lecture.

A control teacher who learned the think-aloud strategy in his credential program along with "a lot of new-age techniques that are not particularly useful; they assume a higher level of maturity than high school" told the interviewer, "I haven't tried think-alouds with the students; I would get mixed results. I might have one good student model it, but I am leery of what it might turn into."

Summary. Differences between intervention and control teachers from several sources of data—surveys, interviews, and assignment samples—indicate that the intervention teachers were more knowledgeable about and more able to integrate the teaching of science reading with science content, to create classrooms characterized by collaborative inquiry and meaning making with science texts, to engage students in the work of text inquiry, and to offer their students tools in the form of comprehension routines and strategies to support their work with science texts. These varied sources of data show intervention teachers implementing instructional approaches consistent with the intervention.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Student Outcomes: Science Understanding, Reading Proficiency, and Engagement in Science Learning

Student OTL survey. To investigate treatment and control group differences on the OTL survey outcomes, we estimated multilevel regression models that included controls for baseline characteristics (randomization strata, race-ethnicity, and whether or not the student reported speaking a non-English language at home). We also examined differences in impacts across self-reported racial-ethnic groups and by whether or not the student reported speaking a language other than English at home.

The student OTL survey results presented in Table 10 corroborated findings from the teacher survey and teacher assignment ratings related to integration of biology and literacy. The results favored the treatment group and were statistically significant for two of the six measures: emphasis on reading in biology—a measure of teacher instruction, guidance, and support for science reading—and student integration of biology and literacy—a measure of student practice of comprehension supporting routines and strategies. In addition, students in treatment schools reported higher levels on the student identity construct than students in control schools ($p = .054$).

No discernable pattern of differences in impacts across racial-ethnic groups was detected, and differences across racial-ethnic groups were not statistically significant. For home language, however, there was limited evidence of differences in impacts across groups. The program impact on reading science was stronger, more positive, and statistically different for students whose home language was not English compared to that for students whose home language was English. Moreover, although not statistically different, the impacts appear to be more robust and consistent for students from families speaking a language other English than for those from English-speaking families, with four out of the six OTL outcomes being statistically significant for other-than-English-background students and one out of the six outcomes being significant for English-background students. For reading in biology, integration of biology and literacy, student identity, and reading science, estimated effect sizes for other-than-English-background students range from 0.34 to 0.43 standard deviations.

State standardized test scores. To examine potential program impacts on student performance in biology and reading comprehension, we examined treatment and control differences in state-mandated criterion-referenced test scores. As described earlier, two types of test score data were collected: linked, longitudinal test score data for students from whom we had obtained parental consent and anonymous, unlinked, cross-sectional data student for students from whom we did not obtain parental consent. To account for treatment and control group nonequivalence in the sample retained, all analyses include controls for student and teacher characteristics measured prior to the intervention, including preintervention measures of student achievement. Table 11 shows the results for both sets of test score data. For

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 10

Student Opportunity to Learn (OTL) Surveys by Treatment and Control Group

OTL Outcome	Treatment	Control	Difference	*p* Value	Difference/*SD*
Reading in biology	3.01	2.85	0.16**	.05	0.30
Non-English	3.01	2.80	0.21**	.03	0.40
English	3.01	2.87	0.14	.11	0.25
Integration of biology and literacy	2.80	2.61	0.19**	.01	0.30
Non-English	2.82	2.55	0.27**	.00	0.43
English	2.78	2.63	0.15**	.04	0.24
Identifying as reader	2.56	2.54	0.02	.81	0.03
Non-English	2.63	2.55	0.08	.49	0.11
English	2.52	2.53	–0.01	.94	–0.01
Student identity	2.84	2.72	0.12*	.05	0.18
Non-English	2.88	2.67	0.21**	.02	0.34
English	2.83	2.74	0.09	.19	0.13
Motivation in class	2.88	2.80	0.09	.16	0.14
Non-English	2.91	2.80	0.11	.23	0.17
English	2.87	2.79	0.08	.26	0.12
Reading science	2.79	2.70	0.09	.33	0.11
Non-English	2.85	2.60	0.25**[a]	.03	0.35
English	2.75	2.74	0.01	.91	0.01

Note. Data are based on regression-adjusted multilevel regression models that include dichotomous indicators for randomization strata, racial-ethnic status indicator variables (Latino, Asian, African American, Other), and primary home language (non-English) as covariates. Effect sizes are calculated by dividing estimates by the pooled standard deviation of the outcome variable. Sample size consists of a maximum of 1,024 students in 42 schools.

[a]Difference in impacts between students whose home language is non-English and English is statistically significant ($p < .05$).

*Treatment-control difference is statistically significant ($p < .10$).

**Treatment-control difference is statistically significant ($p < .05$).

the longitudinal test data, students in treatment schools exhibited similar postintervention levels of performance on the state standardized assessments as their counterparts in control schools. For the cross-sectional data, which was a more representative sample of the students in the study, students in the treatment schools performed better than their counterparts in control schools on all state standardized assessments: English language arts, reading comprehension, and biology.

The effect sizes of 0.23, 0.24, and 0.28 on English language arts, reading comprehension, and biology CST tests give an estimate of the magnitude of the difference between student test scores in the intervention and control groups. A year of reading growth at the high school level has been estimated to produce a magnitude of change of approximately 0.19 (Hill, Bloom, Black, & Lipsey, 2008). This indicates that students in the intervention classes

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 11
State Standardized Test Scores by Treatment and Control Group

Test	Treatment	Control	Difference	*p* Value	Difference/*SD*
2006–2007 (longitudinal)					
ELA CST (std)	–0.02	–0.05	0.03	.74	0.03
Reading comprehension CST (std)	0.05	0.01	0.04	.60	0.04
Biology CST	338.07	330.10	7.97	.27	0.17
2006–2007 (cross-sectional)					
ELA CST (std)	0.10	–0.13	0.23**	.04	0.23
Reading comprehension CST (std)	0.10	–0.14	0.24**	.02	0.24
Biology CST	336.52	322.91	13.61**	.01	0.28

Note. Data are based on regression-adjusted multilevel regression models that include randomization strata, racial-ethnic status indicator variables (Latino, Asian, African American, Other), English language status, and gender as covariates. Longitudinal models also include controls for student English language arts (ELA) and mathematics test scores in the academic year prior to students' enrollment in biology and 2004–2005 teacher-aggregated biology and reading comprehension test scores. Cross-sectional models include controls for 2004–2005 teacher-aggregated biology and reading comprehension test scores. California Standards Test (CST) ELA and reading comprehension scores are standardized (std) to have a mean of 0 and standard deviation of 1, based on the distribution of scores across all students in the analytic sample. Effect sizes are calculated by dividing estimates by the pooled standard deviation of the outcome variable. Longitudinal sample size consists of a maximum of 1,236 students in 45 schools. Cross-sectional sample size consists of a maximum of 5,436 students in 49 schools.

**Treatment-control difference is statistically significant ($p < .05$).

were about 1 year ahead of their counterparts in control classes at the end of the study.

Demographic subgroups. In order to investigate whether the intervention differentially impacted students from groups traditionally unrepresented in the sciences, we also examined program impacts by student racial-ethnic status, English learner status, and gender. For the cross-sectional test data (Table 12), an analysis of scores by demographic group found statistically significant increases in test scores for White and English learner students in the intervention classes. White students and English learners in treatment schools performed better than their counterparts in control schools on all state standardized assessments, with effect sizes ranging from 0.33 to 0.40 for White students and 0.18 to 0.23 for English learners.

Although not statistically significant at conventional levels, positive impacts on ELA and biology test scores were found for Latino students ($p < .10$) as well. Test scores in reading comprehension and biology were higher in intervention schools than control schools for both male and female

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Table 12

State Standardized Test Scores by Subgroup and Treatment and Control Group: Cross-Sectional Results

Test	Treatment	Control	Difference	*p* Value	Difference/*SD*
ELA CST					
Race-ethnicity					
African American	–0.296	–0.370	0.074	.603	0.079
Asian	0.204	0.167	0.037	.785	0.039
Latino	–0.008	–0.184	0.177	.143	0.188
Other	0.163	–0.297	0.460**	.002	0.474
White	0.356	0.019	0.337**	.008	0.328
ELL status					
Non-ELL	0.211	–0.047	0.259	.148	0.255
ELL	–0.079	–0.246	0.167**	.021	0.178
Gender					
Female	0.202	0.017	0.185**	.021	0.190
Male	–0.004	–0.263	0.259	.103	0.256
Reading comprehension					
Race-ethnicity					
African American	–0.274	–0.430	0.156	.246	0.163
Asian	0.222	0.152	0.070	.586	0.078
Latino	0.010	–0.200	0.210*	.074	0.213
Other	0.146	–0.176	0.322**	.041	0.326
White	0.339	–0.055	0.393**	.001	0.396
ELL status					
Non-ELL	0.199	–0.067	0.266*	.081	0.265
ELL	–0.048	–0.249	0.201**	.014	0.206
Gender					
Female	0.209	–0.029	0.238**	.025	0.246
Male	0.000	–0.247	0.247**	.032	0.242
Biology					
Race-ethnicity					
African American	323.511	313.194	10.317	.139	0.242
Asian	344.353	335.873	8.480	.214	0.178
Latino	331.741	319.456	12.285*	.056	0.282
Other	329.033	316.929	12.104	.101	0.260
White	349.437	328.524	20.913**	.001	0.383
ELL status					
Non-ELL	340.051	324.301	15.750	.125	0.298
ELL	331.169	321.717	9.452**	.006	0.227
Gender					
Female	336.848	322.903	13.945**	.011	0.298
Male	336.126	322.867	13.259**	.008	0.259

Note. Data are based on regression-adjusted multilevel regression models. See note for Table 9 for further details. ELA = English language arts; CST = California Standards Test; ELL = English language learner.

*Treatment-control difference is statistically significant ($p < .10$).

**Treatment-control difference is statistically significant ($p < .05$).

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

students. However, no test score differences were found for African American students in treatment versus control schools.

Summary. Thus, our study found some evidence that the intervention—professional development to support implementation of the Reading Apprenticeship instructional framework in high school biology classes—is associated with increases in performance on the state standardized assessments examined. This inference should be tempered by the fact that there is some evidence for selective teacher attrition from the study, whereby teachers who served low-performing students prior to the intervention were more likely to drop out of the study if they were in the control group than if they were in the treatment group. Thus, higher percentages of low-performing students and English learners remained in treatment teachers' classes than in control classes by the end of the study, with the result that the remaining intervention group was skewed toward low-performing students. After controlling for group differences in these preintervention characteristics, the program impacts on test scores were statistically significant. Overall, the pattern of results based on the subgroup analysis of cross-sectional data suggest that the impacts are most consistent and robust for Whites, English learners, and Latino students and equal for females and males in the study.

Discussion

Impact of Reading Apprenticeship Professional Development on Science Instruction

The study reported here demonstrates significant progress in building tools and processes for linking teacher professional development to meaningful classroom change, and from there, to student engagement and achievement, within a scientifically rigorous experimental study design. Multiple measures of teacher implementation reveal a robust corroboration of teacher-level outcomes. Across these measures, teachers in the experimental group demonstrated increased support for science literacy learning, increased use of metacognitive inquiry routines, increased reading comprehension instruction, and increased use of collaborative learning structures. In short, they were more able to integrate science and science literacy learning in classroom instruction than their counterparts in the control group.

Impact of Reading Apprenticeship Professional Development on Student Learning

Student OTL surveys provided evidence that these differences in teaching resulted in learning differences for students, at least in terms of development and support for literacy and student engagement in science reading. Students in intervention classrooms reported more integration of reading

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

and biology, increased confidence in approaching science reading, and in the case of students whose home language is not English, a more robust student identity than students in the control group. In addition, state standardized assessments provided evidence that the intervention—Reading Apprenticeship in biology—is associated with improved performance on state standardized test scores in English language arts, reading comprehension, and biology, with effect sizes of 0.23, 0.24, and 0.28, respectively. These score gains indicate that students in intervention classrooms were a year ahead of their counterparts in control classes by the spring of the school year when tests were given. Moreover, implementation was associated with student performance in expected ways in treatment classrooms. The results of the study thus present a positive picture with regard to the effectiveness of the Reading Apprenticeship framework for integrating academic literacy content with biology coursework and instructional practices.

Limitations of the Study

However, we raise several cautions in interpreting these results. The study utilized a design in which teachers were recruited and randomly assigned to treatment and control groups fully 2 years prior to final data collection. Participating teachers in the treatment group had the opportunity to teach students utilizing what they had learned in the professional development for 2 consecutive academic years. The implication of this sequencing is that treatment teachers had 1 academic year to practice using the framework, which helped ensure that program impacts were assessed after teachers had adequate experience with the framework. However, this aspect of the design required that participating teachers be retained in the study for a lengthy period, yet retaining teachers for such an extended period posed a considerable challenge. We believe that this design, while acknowledging the importance of practice for teachers, exposed the study to greater risk of attrition. Many teachers were reassigned by administrators, dropped out of the study due to changes in districts or schools, or were lost due to changing life circumstances, such as health or maternity leave. Moreover, both treatment and control teachers who participated and were retained in the study exhibited high levels of experience and engagement in other professional development activities targeting science instruction, suggesting that the results generalize to classrooms with experienced science teachers.

Of the 105 teachers randomly assigned to experimental condition, 57 (54%) provided survey, interview, or student test score data at the final data collection point. Analyses of characteristics of schools, teachers, and students who were retained in the sample indicated that treatment schools served higher proportions of students who traditionally do not perform well in school—higher proportions of English learners and Latinos and lower proportions of White students—than did the control schools. Prior to the intervention, retained treatment schools also served greater numbers of lower-performing students than control schools. We think it important to

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

note that these treatment and control group differences in the characteristics of the retained sample may have compromised the experimental design.

Wait-listing the control group for professional development in Reading Apprenticeship was intended to encourage control teachers to remain in the study, since they would ultimately benefit. Even so, attrition was greater in the control group and, as already suggested, particularly in those schools serving the most vulnerable populations of students. Interviews with treatment teachers suggest that the professional development intervention gave them instructional tools to better serve these students. This study thus raises important dilemmas to address in designing randomized controlled studies to be carried out in the ongoing tumult of school systems, particularly when targeting those populations—as did this study—that are most likely to attend the schools experiencing the most administrator and teacher turnover, the toughest challenge to support diverse student needs, and the highest demand from policymakers.

Even with these caveats, the experimental impact estimates reported here are based on models that control for group differences in preintervention student and teacher characteristics. As such, the estimates represent the best estimates of program impacts given data limitations and provide evidence that the Reading Apprenticeship program of professional development can positively impact teaching and learning outcomes in high school biology.

We need to further examine the differences in impacts resulting from the intervention on particular targeted student groups. While English learners, White students, and Latinos benefited from the study, African American students in the intervention schools did not show significantly higher test scores than those in control schools. In preliminary analyses ranking teachers' implementation of Reading Apprenticeship instructional methods, we examined the extent to which level of implementation was related to the types of students served in intervention and control schools. Treatment teachers with low implementation were more likely to teach in schools that exhibited lower levels of school performance. The fact that the most highly impacted schools were associated with the weakest implementation levels for treatment teachers suggests that teachers find it difficult to take on new practices in the context of ongoing challenges in such settings.

The design of this particular study to measure the impact of Reading Apprenticeship professional development limited the possibility of contamination and unwanted "dosage" effects by ensuring that only 9th- or 10th-grade biology teachers in a site could be trained. In most cases, this design meant that implementation teachers were working in isolation. Research indicates that it would be preferable to support teacher implementation efforts by developing a site-based professional community of other science teachers also working to implement Reading Apprenticeship. This would have the additional benefit of increasing the science and literacy instructional "dose" that students would receive. Studies in New Zealand, for

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

instance, have demonstrated that sustained acceleration to close the achievement gap among mainstream and minority populations is possible, with sustained effort at a school site (Lai et al., 2009). A next study warranted by these outcomes might then be to investigate the promise of Reading Apprenticeship professional development as a model of sustained acceleration to transform the hardest-to-reach schools, science teachers, and students. This would require a sustained, longitudinal study involving all science teachers at a school site over multiple years while tracing changes for teachers and cohorts of students as they move through the grade levels. Undertaking such a program of study seems a logical next step for our work.

Conclusions

The Common Core Standards represent an unprecedented set of goals for integrating advanced literacy instruction into disciplinary teaching in the middle and high school grades, and meeting this high level of demand will present school systems with unprecedented challenges. Whatever the outcomes of state adoption processes, efforts to close long-standing achievement gaps and increase the academic proficiencies of graduating youth will of necessity persist as driving forces in education reform. In these efforts, the quality of instruction in discipline areas will be key. If education decision makers are to invest in professional development to build teachers' capacity to meet the high standards envisioned, they will need evidence that professional development can be effective in producing teacher change and improving the quality of instruction and that these changes can result in student achievement gains. These links have too rarely been made. Further, as Desimone (2009) suggests, they will need to know the features of professional development that can be used as indicators of quality and effectiveness. In addition, if content-area teachers are to invest in the work of integrating discipline-based literacy instruction into their curricula, they will need to be shown that such efforts can result in improved student learning of the subject area.

At the outset of this study, we posited that professional development would lead to greater teacher knowledge and practice integrating literacy and science teaching and that these changes would result in increased student engagement and achievement in both literacy and science. This study demonstrates that a generative, inquiry-based model of professional development focused on literacy teaching in an academic content area, such as science, can substantially impact teachers' classroom practices and the resulting opportunities students experience to learn to read and reason with complex science materials and texts. Further, these outcomes indicate that focusing on developing teachers' capacity to provide literacy instruction to support active, intellectual inquiry with science texts can support students' achievement in both reading comprehension skills and biology content learning.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Thus, this study makes a meaningful contribution to the knowledge base about how effective professional development for teachers can result in increased achievement for students, knowledge that can be used to inform education reform. Increasingly, proposed strategies for improving educational outcomes for underachieving students focus on making structural changes to schools, increasing teacher and principal accountability, and redistributing effective teachers through incentives for working in the most challenging schools. In this context, the outcomes of this study are highly significant, demonstrating that it is possible to improve outcomes for students by investing as well in building existing teachers' capacity through well-designed professional development interventions of this kind.

We believe this study also demonstrates the promise of taking a disciplinary approach to literacy instruction, showing through use of rigorous research methodologies that it is possible to improve the instructional quality of science teaching at the high school level through professional development focused on literacy in science learning and that these changes can result in improved engagement and learning for students. Further, at a time when struggling secondary students are increasingly removed from content-area learning to remediate their literacy skills, this study makes a contribution of great potential practical import, since integrating literacy and science instruction not only would mitigate the unintended consequences of restricting students' access to vital content area learning but could result in substantial cost savings to districts and schools. The results of this study indicate that integrated literacy instruction can support, rather than supplant, science learning for students and, conversely, that an instructional focus on developing students' reading proficiencies in specific disciplines, like science, can meaningfully improve students' reading comprehension and literacy, more generally.

Notes

This work was supported by Grant No. 0440379 from the Research on Learning in Education (ROLE) Program of the National Science Foundation to Steven Schneider, principal investigator (PI); Cynthia Greenleaf, co-PI; and Joan Herman, co-PI.

[1]Because one of the PIs for this study is the developer of the Reading Apprenticeship framework and has carried out a program of research and development focused on effective professional development for Reading Apprenticeship, the research team carefully delineated roles to avoid the possibility or the perception of bias in the study of this intervention. The primary role of WestEd's Math and Science program and the Strategic Literacy Initiative project in the Teacher Quality program of WestEd was to provide content expertise in science, literacy, and professional development to inform the intervention and instrumentation of the study. Strategic Literacy Initiative staff were further divided into two formally isolated teams with research versus professional development responsibilities. The primary role of the Center for Research on Evaluation, Standards, and Student Testing (CRESST) at the University of California, Los Angeles, was to develop, field test, and analyze measures of the nature and degree of literacy instruction in biology classrooms, including

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

teacher surveys and teacher assignments. All scoring of these measures was carried out by CRESST and nonparticipant teachers whom CRESST staff trained to score. As content experts, the Strategic Literacy Initiative staff were involved in the scoring of teacher interviews, alone. The Math and Science staff, with Research and Evaluation experts at the Los Alamitos office of WestEd, kept locked data files and codes identifying teachers and students, to which only they had access.

[2]Because odd numbers of schools were eligible for randomization in four of the six waves, uneven numbers of schools were assigned to intervention conditions (43 treatment and 40 control schools) due to chance factors alone.

[3]Readers may contact the first author for a copy of the full survey.

References

Abell, S. A. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea? *International Journal of Science Education*, *30*(10), 1405–1416.

ACT. (2006). *Reading between the lines: What the ACT reveals about college readiness in reading*. Retrieved from http://www.act.org/path/policy/pdf/reading_report.pdf

Allington, R. L. (1991). Children who find learning to read difficult: School responses to diversity. In E. H. Hiebert (Ed.), *Literacy for a diverse society: Perspectives, practices, and policies* (pp. 237–252). New York, NY: Teachers College Press.

Allington, R. L., & McGill-Franzen, A. (1989). School response to reading failure: Chapter 1 and special education students in Grades 2, 4, and 8. *Elementary School Journal*, *89*, 529–542.

Alvermann, D., & Moore, D. (1991). Secondary school reading. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 951–983). New York, NY: Longman.

American Association for the Advancement of Science. (1993). *Project 2061: Benchmarks for science literacy*. New York, NY: Oxford University Press.

Aschbacher, P. R. (1999). Developing indicators of classroom practice to monitor and support school reform (CSE Tech. Rep. No. 513). Los Angeles: University of California, Los Angeles/Center for Research on Evaluation, Standards, and Student Testing.

Baker, L. (1991). Metacognition, reading, and science education. In C. Santa & D. Alvermann (Eds.), *Science learning: Processes and applications* (pp. 2–13). Newark, DE: International Reading Association.

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

Barton, P. (2003). *Parsing the achievement gap: Baselines for tracking progress*. Policy information report of the Educational Testing Service. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/research/pic

Baumann, J. F., & Duffy, A. M. (1997). *Engaged reading for pleasure and learning: A report from the National Reading Research Center*. Athens, GA: National Reading Research Center.

Bayer, A. S. (1990). *Collaborative apprenticeship learning: Language and thinking across the curriculum, K–12*. Mountain View, CA: Mayfield.

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

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15.

Boston, M. D., & Smith, M. S. (2009). Transforming secondary mathematics teaching: Increasing the cognitive demands of instructional tasks used in teachers' classrooms. *Journal for Research in Mathematics Education*, *40*(2), 119–156.

Bransford, J., Derry, S., Berliner, D., & Hammerness, K. (2005). Theories of learning and their role in teaching. In L. Darling- Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world* (pp. 40–87). San Francisco, CA: Jossey-Bass.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.

Brown, J. S., Collins, A., & Newman, S. (1989). The new cognitive apprenticeship: Teaching the craft of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum.

Bybee, R. W. (1995). Achieving scientific literacy: Using the National Science Education standards to provide equal opportunities for all students to learn science. *Science Teacher*, *62*(7), 28–33.

Bybee, R. W. (1997). *Achieving scientific literacy: From purposes to practices*. Portsmouth, NH: Heinemann.

Calfee, R., & Miller, R. G. (2004, April). *Separate no more: A theoretical and practical basis for using embedded reading and writing instruction for expository text*. Paper distributed at the annual meeting of the American Educational Research Association, San Diego, CA.

California Department of Education. (2000). *Construction of California's 1999 School Characteristics Index and similar schools ranks*. Sacramento, CA: Author, Office of Policy and Evaluation.

Camburn, E., Rowan, B., & Taylor, J.E. (2003). Distributed leadership in schools: The case of elementary schools adopting comprehensive school reform models. *Educational Evaluation and Policy Analysis*, *25*, 347–373.

Carnegie Council on Advancing Adolescent Literacy. (2010). *Time to act: An agenda for advancing adolescent literacy for college and career success*. New York, NY: Carnegie Corporation of New York.

Center on Education Policy. (2006*). Ten big effects of the No Child Left Behind act on public schools*. Washington, DC: Author.

Cervetti, G., Pearson, P. D., Bravo, M. A., & Barber, J. (2006). Reading and writing in the service of inquiry-based science. In R. Douglas, M. Klentschy, & K. Worth (Eds.), *Linking science and literacy in the K–8 classroom* (pp. 221–244). Arlington, VA: National Science Teachers Association Press.

Clare, L. (2000). *Using teachers' assignments as an indicator of classroom practice* (CSE Tech. Rep. No. 532). Los Angeles: University of California, Los Angeles/ Center for Research on Evaluation, Standards, and Student Testing.

Clare, L., & Aschbacher, P. (2001). Exploring the technical quality of using assignments and student work as indicators of classroom practice. *Educational Assessment*, *7*, 39–59.

Clark, K. & Borko, H. (2004). Establishing a professional learning community among middle school mathematics teachers. In M. J. Hoines & A. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 223–230). Bergen, Norway: Bergen University College.

Cohen, D., & Ball, D. (2000). *Instructional innovation: Reconsidering the story*. The Study of Instructional Improvement working paper. Retrieved from http://www.sii.soe.umich.edu/documents/InstructionalInnovation.pdf

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Conley, M. W. (2008). Cognitive strategy instruction for adolescents: What we know about the promise, what we don't know about the potential. *Harvard Educational Review*, *78*(1), 84–108.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues in field settings*. Chicago, IL: Rand-McNally.

Cunningham, A., & Stanovich, K. (1998, Spring/Summer). What reading does for the mind. *American Educator*, pp. 8–15.

Darling-Hammond, L., & Sykes, G. (2003). Wanted: A national teacher supply policy for education: The right way to meet the "highly qualified teacher" challenge. *Education Policy Analysis Archives*, *11*(33), 1–57.

Darling-Hammond, L., Wei, R. C., 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.

Davidson, J., & Koppenhaver, D. (1993). *Adolescent literacy: What works and why* (2nd ed.). New York, NY: Garland.

Delpit, L. D. (1995). *Other people's children: Cultural conflict in the classroom*. New York, NY: New Press.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199.

Desimone, L. M., Garet, M., Birman, B., Porter, A., Yoon, K. S. (2002). How do district management and implementation strategies relate to the quality of the professional development that districts provide to teachers? *Teachers College Record*, *104*(7), 1265–1312.

Digisi, L. L., & Willett, J. B. (1995). What high school biology teachers say about their textbook use: A descriptive study. *Journal of Research in Science Teaching*, *32*, 123–142.

Donahue, P. L., Voelkl, K. E., Campbell, J. R., & Mazzeo, J. (1999). *The NAEP 1998 reading report card for the nation and the states*. Washington, DC: National Center for Education Statistics.

Donovan, S. M., Bransford, J. D., & Pellegrino, J. W. (Eds.). (1999). *How people learn: Bridging research and practice*. Washington, DC: National Academy Press.

Duffy, G. G., Roehler, L. R., Meloth, M. S., Vavrus, L. G., Book, C., Putnam, J. & Wesselman, R. (1986). The relationship between explicit verbal explanations during reading skill instruction and student awareness and achievement: A study of reading teacher effects. *Reading Research Quarterly*, *21*(3), 237–252.

Duke, N. (2000). 3.6 minutes per day: The scarcity of informational texts in first grade. *Reading Research Quarterly*, *35*(2), 202–224.

Durkin, D. (1978). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly*, *14*(4), 481–533.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (Eds.). (2006). *Taking science to school: Learning and teaching science in Grades K–8*. Washington, DC: National Academies Press.

Dweck, C., & Molden, D. (2005). Self-theories: Their impact on competence motivation and acquisition. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 122–140). New York, NY: Guilford.

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V. R., & Empson, S. B. (1996). A longitudinal study of learning to use children's thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27*(4), 403–434.

Fielding, L. G., & Pearson, D. P. (1994). Reading comprehension: What works. *Educational Leadership*, *51*, 62-68.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Fishman, B. J., Marx, R. W., Best, S. & Tal, R. T. (2003). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, *19*, 643–658.

Freedman, S. W., Flower, L., Hull, G., & Hayes, J. R. (1995). *Ten years of research: Achievements of the National Center for the Study of Writing and Literacy* (Tech. Rep. No. 1-C). Berkeley, CA: National Center for the Study of Writing.

Fullan, M. (1991). *The new meaning of educational change*. New York, NY: Teachers College Press.

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*, 915–945.

Gee, J. (1999). Critical issues: Reading and the new literacy studies: Reframing the National Academy of Science report on reading. *Journal of Literacy Research, 31*(3), 355–374.

Goldstein, H. (1987). *Multilevel models in educational and social research*. London, UK: Griffin; New York: Oxford University Press.

Greenleaf, C. (2006, February). Fostering metacognitive conversation in professional communities and subject-area classrooms. *Adolescent Literacy in Perspective*, pp. 2–5. Retrieved from http://www.ohiorc.org/adlit

Greenleaf, C., Brown, W., & Litman, C. (2004). Apprenticing urban youth to science literacy. In D. Strickland & D. Alvermann (Eds.), *Bridging the gap: Improving literacy learning for preadolescent and adolescent learners in Grades 4–12*. Newark, NJ: International Reading Association.

Greenleaf, C., & Hinchman, K. (2009). Reimagining our inexperienced adolescent readers: From struggling, striving, marginalized, and reluctant to thriving. *Journal of Adolescent and Adult Literacy*, *53*(1), 4–13.

Greenleaf, C., Hull, G., & Reilly, B. (1994). Learning from our diverse students: Helping teachers rethink problematic teaching and learning situations. *Teaching and Teacher Education*, *10*(5), 521–541.

Greenleaf, C., & Katz, M. (2004). Ever newer ways to mean: Authoring pedagogical change in secondary subject-area classrooms. In S. W. Freedman & A. F. Ball (Eds.), *Bakhtinian perspectives on language, literacy, and learning* (pp. 172–202). Cambridge, UK: Cambridge University Press.

Greenleaf, C., & Schoenbach, R. (2001). *Close readings: A study of key issues in the use of literacy learning cases for the professional development of secondary teachers*. Final report to the Spencer and MacArthur Foundations Professional Development Research and Documentation Program. Oakland, CA: Strategic Literacy Initiative, WestEd.

Greenleaf, C., & Schoenbach, R. (2004). Building capacity for the responsive teaching of reading in the academic disciplines: Strategic inquiry designs for middle and high school teachers' professional development. In D. S. Strickland & M. L. Kamil (Eds.), *Improving reading achievement through professional development* (pp. 85–96). Norwood, MA: Christopher-Gordon Publishers, Inc.

Greenleaf, C., Schoenbach, R., Cziko, C., & Mueller, F. (2001). Apprenticing adolescents to academic literacy. *Harvard Educational Review, 71*(1), 79–129.

Grigg, W. S., Lauko, M. A., & Brockway, D. M. (2006). *The nation's report card: Science 2005* (NCES 2006-466). Washington, DC: Government Printing Office.

Grossman, P., Wineburg, S., & Woolworth, S. (2001). Toward a theory of teacher community. *Teachers College Record*, *103*, 942–1012.

Groves, F. (1995). Science vocabulary load of selected secondary science textbooks. *School Science and Mathematics, 95* (5), pp. 231–235

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Guskey, T. R. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, *8*(3/4), 381–391.

Guskey, T. R., & Huberman, M. (1996). *Professional development in education: New paradigms and practices*. Columbia, NY: Teachers College Press.

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

Guthrie, J. T., & Greaney, V. (1991). Literacy acts. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 68–96). New York, NY: Longman.

Guthrie, J. T., McGough, K., Bennett, L., & Rice, M. E. (1996). Concept-oriented reading instruction: An integrated curriculum to develop motivations and strategies for reading. In L. Baker, P. Afflerbach, & D. Reinking (Eds.), *Developing engaged readers in school and home communities* (pp. 165–190). Mahwah, NJ: Lawrence Erlbaum.

Guthrie, J. T., Schafer, W. D., & Hutchinson, S. R. (1991). Relations of document literacy and prose literacy to occupational and societal characteristics of young Black and White adults. *Reading Research Quarterly*, *26*, 30–48.

Hargreaves, A. (2003). *Teaching in the knowledge society: Education in the age of insecurity*. New York, NY: Teachers College Press.

Hart, J., & Lee, O. (2003). Teacher professional development to improve the science and literacy achievement of English language learners. *Bilingual Research Journal*, *27*(3), 475–501.

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership*, *58*(6), 6–11. Retrieved from http://www.ascd.org/readingroom/edlead/0103/haycock.html

Heller, R., & Greenleaf, C. (2007). *Literacy instruction in the content areas: Getting to the core of middle and high school improvement*. Washington, DC: Alliance for Excellent Education.

Hiebert, E. (1991). *Literacy for a diverse society: Perspectives, policies, and practices*. New York, NY: Teachers College Press.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177.

Huberman, M. (1995). Networks that alter teaching: Conceptualizations, exchanges and experiments. *Teachers and Teaching: Theory and Practice*, *1*(2), 193–211.

Huffman, D., Thomas, K., & Lawrenz, F. (2003). Relationship between professional development, teachers' instructional practices, and the achievement of students in science and mathematics. *School Science and Mathematics*, *103*(8), 378–387.

Hull, G., & Rose, M. (1989). Rethinking remediation: Toward a social-cognitive understanding of problematic reading and writing. *Written Communication*, *6*(2), 139–154.

Hynd, C. (1998). *Learning from text across conceptual domains*. Mahwah, NJ: Lawrence Erlbaum.

Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices. A practice guide* (NCEE 2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc

Kazemi, E., & Franke, M. L. (2004). Teacher learning in mathematics: Using student work to promote collective inquiry. *Journal of Mathematics Teacher Education*, *7*, 203–235.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Kennedy, M. (1998). *Form and substance in inservice teacher education* (Research Monograph No. 13). Madison: University of Wisconsin-Madison, National Institute for Science Education.

Killion, J. (1998). Scaling the elusive summit. *Journal of Staff Development*, *19*(4), 12–16.

Klingner, J. K. (2004). The science of professional development. *Journal of Learning Disabilities*, *37*(3), 248–255.

Knapp, M. S. (1995). *Teaching for meaning in high-poverty classrooms*. New York, NY: Teachers College Press.

Kucan, L., & Beck, I. L. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, *67*(3), 271–299.

Lai, M. K., McNaughton, S., Amituanai-Toloa, M., Turner, R., & Hsiao, S. (2009). Sustained acceleration of achievement in reading comprehension: The New Zealand experience. *Reading Research Quarterly*, *44*(1), 30–56.

Langer, J. (2001). Beating the odds: Teaching middle and high school students to read and write well. *American Educational Research Journal*, *38*, 837–880.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Lee, C. D. (1995). A culturally based cognitive apprenticeship: Teaching African American high school students skills in literary interpretation. *Reading Research Quarterly*, *30*, 608–630.

Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy*. New York, NY: Carnegie Corporation of New York.

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

Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex.

Lemke, J. L. (2006). Towards critical multimedia literacy: Technology, research, and politics. In M. G. McKenna, L. D. Labbo, R. D. Keiffer, & D. Reinking (Eds.), *International handbook of literacy and technology* (Vol. 2, pp. 3–14). Hillsdale, NJ: Lawrence Erlbaum.

Litman, C., & Greenleaf, C. (2008). Traveling together over difficult ground: Negotiating success with a profoundly inexperienced reader in an introduction to chemistry class. In K. A. Hinchman & H. K. Sheridan-Thomas (Eds.), *Best practices in adolescent literacy instruction* (pp. 275–296). New York, NY: Guilford.

Little, J. W. (2001, September). *Inside teacher community: Representations of classroom practice*. Paper presented as an invited address at the biannual conference of the International Study Association on Teachers and Teaching, Faro, Portugal.

Little, J. W. (2002). Locating learning in teachers' communities of practice: Opening up problems of analysis in records of everyday work. *Teaching and Teaching Education*, *18*, 917–946.

Loucks-Horsley, S., Love, N., Styles, K. E., Mundry, S., & Hewson, P. W. (2003). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science. *School Science and Mathematics*, *99*(5), 258–271.

Manzo, A. (1969). ReQuest: A method for improving reading comprehension through reciprocal questioning. *Journal of Reading*, *12*(3), 123–126.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Matsumura, L. C. (2003). Teachers, assignments and student work: Opening a window on classroom practice (CSE Tech. Rep. No. 602). Los Angeles: University of California, Los Angeles/Center for Research on Evaluation, Standards, and Student Testing.

McLaughlin, M. W., & Talbert, J. E. (2001). *Professional communities and the work of high school teaching*. Chicago, IL: University of Chicago Press.

McMahon, M., & McCormack, B. (1998). To think and act like a scientist: Learning disciplinary knowledge. In C. Hynd (Ed.), *Learning from text across conceptual domains* (pp. 227–262). Mahwah, NJ: Lawrence Erlbaum.

McMurrer, J. (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Washington, DC: Center on Educational Policy.

Miller, R. G. (2004). *Creating thoughtful and expressive teachers and students using read-write activities: Final report to NSF*. IEA-ETS Research Institute.

Moje, E. B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent and Adult Literacy, 52*(2), 96–107.

Moje, E. B., Ciechanowski, K. M., Kramer, K., Ellis, L., Carrilo, R., & Collazo, T. (2004). Working toward third space in content area literacy: An examination of everyday funds of knowledge and discourse. *Reading Research Quarterly*, *39*(1), 38–70.

Moje, E. B., Dillon, D. R., & O'Brien, D. G. (2000). Re-examining the roles of the learner, the text, and the context in secondary literacy. *Journal of Educational Research*, *93*, 165–180.

Moje, E. B., Overby, M., Tysvaer, N., & Morris, K. (2008). The complex world of adolescent literacy: Myths, motivations, and mysteries. *Harvard Educational Review*, *78*(1), 107–154.

Mullis, I. V., Dossey, A., Campbell, J. R., Gentile, C. A., O'Sullivan, C., & Latham, A. S. (1994). *NAEP 1992 trends in academic progress* (Report No. 23-TR01). Washington, DC: Government Printing Office.

Murray, D. M. (1998). *Design and analysis of group randomized trials*. New York, NY: Oxford University Press.

National Assessment of Educational Progress. (2006). *The nation's report card: Science 2005*. Washington, DC: Author. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006466

National Assessment of Educational Progress. (2007). *The nation's report card: Reading 2007*. Washington, DC: Author. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007496

National Assessment of Educational Progress. (2009). NAEP 2008 trends in academic progress (NCES 2009-469). Washington, DC: National Center for Education Statistics.

National Center for Education Statistics. (2007). *Highlights from PISA 2006: Performance of U.S. 15-year-old students in science and mathematics literacy in an international context* (NCES 2008–016). Washington, DC: Author.

National Committee on Science Education Standards and Assessment. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Council of Chief State School Officers & National Governors Association. (2010). *Common core standards for English Language Arts and literacy in history/social studies and science*. Retrieved from http://www.corestandards.org/Standards/K12/

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Department of Health and Human Services, National Institute of Child Health and Human Development.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

National Research Council. (1990). *Fulfilling the promise: Biology education in the nation's schools*. Washington, DC: National Academy Press.

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, *66*, 60–92.

Norris, S. P., & Phillips, L. M. (2003). How literacy in its fundamental sense is central to scientific literacy. *Science Education*, *87*, 224–240.

Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). New Haven, CT: Yale University Press.

Osborne, J. (2002). Science without literacy: A ship without a sail? *Cambridge Journal of Education*, *32*(2), 203–218.

Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*, 463–466.

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, *62*, 307–332.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*(2), 117–175.

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly*, *40*(2), 184–202.

Pearson, P. D. (1996). Reclaiming the center. In M. Graves, P. van den Broek, & B. M. Taylor (Eds.), *The first R: Every child's right to read* (pp. 259–274). New York, NY: Teacher's College Press.

Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science*, *328*, 459–463.

Phelps, R., Graham, A., & Kerr, B. (2004). Teachers and ICT: Exploring a metacognitive approach to professional development. *Australasian Journal of Educational Technology*, *20*(1), 49–68.

Porter, A., Garet, M., Desimone, L., & Birman, B. (2003). Providing effective professional development: Lessons from the Eisenhower program. *Science Educator*, *12*(1), 23–40.

Pressley, M. (1998). *Reading instruction that works: The case for balanced teaching*. New York, NY: Guilford Press.

Umphrey, J. (2009). Toward 21st century supports: An interview with Linda Darling-Hammond. *Principal Leadership*, *10*(1), 18–21.

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

Reed, D. K. (2009). A synthesis of professional development on the implementation of literacy strategies for middle school content area teachers. *Research in Middle Level Education Online*, *32*(10), 1–12.

Rentner, D. S., Scott, C., Kober, N., Chudowsky, N., Chudowsky, V., Joftus, S., & Zabala, D. (2006). *From the capital to the classroom: Year 4 of the No Child Left Behind Act*. Washington, DC: Center on Education Policy.

Richardson, V. (Ed.). (1994). *Teacher change and the staff development process*. New York, NY: Teachers College Press.

Richardson, V. (2003). The dilemmas of professional development. *Phi Delta Kappan*, *84*(5), 401–406.

Romance, N. R., & Vitale, M. R. (2001). Implementing an in-depth expanded science model in elementary schools: Multi-year findings, research issues, and policy implications. *International Journal of Science Education, 23*(4), 373–404.

Romance, N. R., & Vitale, M. R. (2008, November 12). *Perspectives for improving school instruction and learning: An interdisciplinary model for integrating science and reading in Grades K–5*. Paper presented at the Committee on Education Workshop on Education lecture series, Chicago, IL. Retrieved from

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

http://www.scienceideas.org/pubs-pres/articles-instructionalmodel/Improving-School-Learning-Instruction.pdf

Roth, K. (1991). Reading science texts for conceptual change. In C. Santa & D. Alvermann (Eds.), *Science learning: Processes and applications* (pp. 48–63). Newark, DE: International Reading Association.

Rutherford, F. J., & Ahlgren, A. (1990). *Science for all Americans*. New York, NY: Oxford University Press.

Rycik, J., & Irvin, J. (Eds.). (2001). *What adolescents deserve: A commitment to students' literacy learning*. Newark, DE: International Reading Association.

Saul, W. E. (Ed.). (2004). *Crossing borders in literacy and science instruction: Perspectives on theory and practice*. Newark, DE: International Reading Association.

Schoenbach, R., & Greenleaf, C. (2009). Fostering adolescents' engaged academic literacy. In L. Christenbury, R. Bomer, & P. Smagorinsky (Eds.), *Handbook of adolescent literacy research*, (pp. 98–112). New York, NY: Guilford Press.

Schoenbach, R., Greenleaf, C., Cziko, C., & Hurwitz, L. (1999). *Reading for understanding: A guide to improving reading in middle and high school classrooms*. San Francisco, CA: Jossey-Bass.

Scott, J. (1993). *Science and language links: Classroom implications*. Portsmouth, NH: Heinemann.

Scribner, S., & Cole, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*(1), 40–59.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*, 4–14.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*, 114–135.

Sizer, T. (1992). *Horace's compromise: The dilemma of the American high school*. Boston, MA: Houghton Mifflin.

Snipes, J., & Horwitz, A. (2008). *Advancing adolescent literacy in urban schools*. Washington, DC: Council of the Great City Schools.

Snow, C. (2010). Academic language and the challenge of reading for learning about science. *Science*, *328*, 450–452.

Snow, C. (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Arlington: RAND. Retrieved from http://www.rand.org/publications/MR/MR1465/

Somers, M.-A., Corrin, W., Sepanik, S., Salinger, T., Levin, J., & Zmach, C. (2010). *The Enhanced Reading Opportunities Study final report: The impact of supplementary literacy courses for struggling ninth-grade readers* (NCEE 2010-4021). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Sparks, D., & Hirsh, S. (1997). *A new vision for staff development: Results-driven education, systems thinking, and constructivism*. Oxford, OH: National Staff Development Council.

Sparks, D. (2004). The looming danger of a two-tiered professional development system. *Phi Delta Kappan*, *86*, 304–306.

Spillane, J. P. (1999). External reform initiatives and teachers' efforts to reconstruct their practice: The mediating role of teachers' zones of enactment. *Journal of Curriculum Studies*, *31*(2), 143–175.

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360–407.

Strickland, D. S., & Kamil, M. L. (Eds.). (2004). *Improving reading achievement through professional development*. Norwood, MA: Christopher-Gordon.

Supovitz, J. A. (2001). Translating teaching practice into improved student performance. In S. H. Fuhrman (Ed.), *From the Capitol to the classroom: Standards-based reform in the states*. The 100th yearbook of the National Society for the Study of Education, Part 2 (pp. 81–98). Chicago, IL: University of Chicago Press.

Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. *Educational Policy*, *14*(3), 331–356.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, *37*, 963–980.

Their, M., with Davis, B. (2002). *The new science literacy: Using language skills to help students learn science*. Portsmouth, NH: Heinemann.

van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, *328*, 453–456.

vanDriel, J. H., Beijaard, D., & Verloop, N. (2001). Professional development and reform in science education: The role of teachers' practical knowledge. *Journal of Research in Science Teaching*, *38*(2), 137–158.

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.

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of K-12 mathematics and science education in the United States*. Chapel Hill, NC: Horizon Research.

Whigham, M., Andre, T., & Yang, E. (2000, April). *Elementary and secondary teachers' beliefs about and instructional emphasis on the National Mathematics Education and Science Education standards*. Paper presented at the annual meeting of the National Association of Research in Science Teaching, New Orleans, LA.

Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, *24*, 173–209.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues and Answers Report REL 2007 No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Education Laboratory Southwest.

Zech, L. K., Gause-Vega, C. L., Bray, M. H., Secules, T., & Goldman, S. R. (2000). Content-based collaborative inquiry: A professional development model for sustaining educational reform. *Educational Psychologist*, *35*(3), 207–217.

Manuscript received March 6, 2010
Final revision received July 13, 2010
Accepted August 19, 2010

Downloaded from http://aerj.aera.net at University of Canberra on June 21, 2016