INTEGRATING LITERACY AND SCIENCE INSTRUCTION IN HIGH SCHOOL BIOLOGY: IMPACT ON TEACHER PRACTICE, STUDENT ENGAGEMENT, AND STUDENT ACHIEVEMENT

Cynthia Greenleaf*, Thomas Hanson*, Joan Herman**, Cindy Litman*, Sarah Madden*
Rachel Rosen*, Christy Boscardin**, Steve Schneider*, and David Silver**

December, 2009
Final report to the National Science Foundation

This work was supported by Grant #0440379
from the Research on Learning in Education (ROLE) Program
of the National Science Foundation
to Steven Schneider, PI
Cynthia Greenleaf, Co-PI,
and Joan Herman, Co-PI

* WestEd
300 Lakeside Drive, 25th Floor
Oakland, CA 94612-3534

** University of California-Los Angeles
National Center for Research on
Evaluation, Standards, & Student Testing
300 Charles E. Young Drive North
GSE&IS Bldg. 3rd Flr./Mailbox 951522
Los Angeles, CA 90095-1522

EXECUTIVE SUMMARY

This report describes findings from a research study examining the effects of integrating academic literacy instruction with biology coursework on teacher instructional practices and student achievement in science and reading. The research builds on a well-tested approach to literacy instruction, Reading Apprenticeship (RA) (Greenleaf, et al., 2001), which integrates metacognitive inquiry into ongoing content area instruction to make explicit the tacit reasoning processes, strategies, and discourse rules that shape successful readers' and writers' work in the discipline. The instructional model draws together research-based practices in reading instruction, including methods of engaging students in extensive reading; integrating explicit teaching of comprehension strategies; establishing relevance and making personal connections to reading materials and curriculum activities; identifying and using a variety of text structures to support comprehension; and supporting collaborative sense-making activities with written materials. The central dynamic of this instructional model is routine metacognitive conversation; that is, talking about the reasoning and problem solving processes that accompany reading as students carry out learning tasks in the science curriculum.

A highly designed program of professional development in Reading Apprenticeship was the intervention explored in this efficacy study (Greenleaf & Schoenbach, 2004). The professional development curriculum was designed to involve teachers in inquiry into their own science literacy practices and in close analysis of text and task demands, as well as inquiries into student literacy performances through videotapes of class and individual student reading activities, written case studies, and ongoing student assessment. Professional development was also designed to model target instructional approaches, engaging teachers in practicing metacognitive routines, modeling reading and reasoning processes, conducting small group work, engaging and supporting students in extended reading opportunities, and facilitating discussions that focus on how and why to read science as well as the science content of science texts. These instructional approaches were tightly integrated with core units of study in biology to illustrate the integration of literacy and science learning.

A group randomized, experimental design was used to assess program impacts. From an initial recruited sample of 105 teachers in 83 high schools, approximately 87 teachers (and their students) from 70 pair-matched high schools participated in the study in year one – half of which were assigned to an immediate professional development group and half of which were wait-listed to receive professional development subsequent to study participation. In year two, 60 teachers from 48 high schools remained in the study. The study relied on multiple measures of teacher implementation and student engagement and learning: (a) pre- and post-intervention surveys of teachers reporting their instructional practices and beliefs about reading, student learning, and student diversity; (b) teacher interview data about instructional practices, beliefs, and student engagement in literacy learning opportunities; (c) teacher practices as reflected by teacher assignments; (d) student opportunity to learn (OTL) surveys; (e) student performance on Integrated Learning Assessments (ILA), and (f) standardized test results derived from the California Standards Tests in English language arts, reading comprehension, and biology. Hierarchical linear modeling procedures were used to estimate program impacts on teacher and student outcomes.

The multiple measures of teacher implementation give us a robust corroboration of teacher level outcomes. 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. In particular, teachers in schools randomly assigned to the Reading Apprenticeship treatment group exhibited significantly higher implementation on 7 of the 14 constructs assessed by the teacher surveys, 4 of the 6 constructs assessed via teacher interviews (*Teacher Support, Metacognitive Inquiry, Specific Tools and Resources*, and *Collaboration*), 2 domains assessed by teacher assignment ratings (*Support for Cognitive Challenge in Biology* and in *Literacy*), and 2 of 6 constructs assessed by the student opportunity to learn surveys (*Reading in Biology, Integration of Biology and Literacy*), compared to the randomly assigned control group.

Student Opportunity to Learn (OTL) surveys and Integrated Learning Assessments (ILA) provided evidence, or leading indicators, that these differences in teaching resulted in learning differences for students, particularly in the ways students engage in science reading tasks. OTL surveys also show some evidence of disproportionate program benefit for students whose home language is not English.

To examine potential program impacts on student performance in state-mandated criterion-referenced test scores, two types of test score data were collected — linked, longitudinal test score data for students for whom we had obtained parental consent; and anonymous, unlinked, cross-sectional data for students for whom we did not obtain parental consent. To enhance the precision of the impact estimates and to account for potential differences in pre-intervention characteristics between groups, the test score analyses controlled for student and teacher characteristics. For the longitudinal test data, students in treatment schools exhibited similar levels of performance on the state standardized assessments as their counterparts in control schools. For the cross-sectional data, after controlling for existing treatment/control school differences in teacher characteristics among the sample retained in year 2, 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. Thus, there is 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. An analysis of scores by demographic group found statistically significant increases in test scores for white, Latino, and English learner students in the intervention classes.

Integrating Literacy and Science Instruction in High School Biology: Impact on Teacher Practice, Student Engagement, and Student Achievement

INTRODUCTION

Our democracy and future economic well-being depend on a literate populace, capable of fully participating in the demands of the 21st century (Rutherford & Algren, 1990). Yet NAEP results indicate that most young people lack the capabilities to successfully engage in the higher-level literacy, scientific understandings and inquiry skills needed for an information generating and transforming economy (NAEP, 2006, 2007). A persistent achievement gap persists between mainstream populations and those who are outside of that mainstream (Gee, 1999). As a result, low performing students often are placed into skill-based classrooms with the goal of increasing their literacy proficiency, yet the skill-based instruction they receive may perpetuate low literacy achievement rather than accelerate literacy growth (e.g. Hull & Rose, 1989; Knapp & Turnbull, 1991). Further, the reading and intervention programs struggling students receive to address reading problems increasingly means that these students lose opportunities to engage in other academic subjects, particularly science (Herman, 2008; McMurrer, 2007).

This study advances the idea that we must think strategically about the integration of development across subject matter domains if we expect to develop students' multiple capacities, particularly those from groups who have been historically underrepresented in the sciences. Science classrooms conceivably can contribute opportunities for students to acquire greater literacy proficiency, and greater literacy proficiency also is essential to students' acquisition of deep scientific understandings and inquiry skills. A key premise of this initiative is that science inquiry and literacy practices share important properties that make the integration of literacy and science particularly powerful (Cervetti, Pearson, Bravo, & Barber, 2006; Greenleaf, Brown & Litman, 2004; McMahon & McCormack, 1998). Participation in investigation-oriented science relies on various kinds of sophisticated literacy skills - the ability to access scientific terminology, interpret arrays of data, comprehend scientific texts, and read and write scientific explanations (Norris & Phillips, 2003; Osborne, 2002).

National policy responses to persistent achievement inequities increasingly target improving teacher quality through national funding priorities (www.ed.gov), but these policy discussions frequently focus on the need to recruit and retain better prepared teachers and to distribute them more equitably to schools serving low achieving student populations, rather than on the parallel endeavors needed to identify effective means of *building* teacher quality (but see National Education Association, link). This study advances our knowledge of the role carefully designed professional development can play in developing teachers' knowledge and practice.

This research study examines the effects of a promising professional development program, Reading Apprenticeship, for building teachers' ability to integrate disciplinary literacy practices into science teaching in high school biology classes, exploring the resulting changes in teacher knowledge and skills, instructional practices, and student achievement in science and reading. The Reading Apprenticeship instructional model draws on research in sociocognitive apprenticeship research in reading and literacy (Lee, 2005), integrating metacognitive inquiry into ongoing content area instruction to make explicit the tacit reasoning processes, strategies, and discourse rules that shape successful readers' and writers' work in the discipline. The central dynamic of this instructional model is routine metacognitive conversation; that is, talking about the reasoning and problem solving processes that accompany reading. The professional development curriculum involves teachers in practicing metacognitive routines, modeling reading and reasoning processes, conducting small group work, engaging and supporting students in extended reading opportunities, and facilitating class discussions that focus on how to read science and why people read science materials in the ways they do as well as the science content of science texts. These instructional approaches are tightly integrated with core units of study in biology to illustrate such integration.

THEORETICAL BACKGROUND AND RELEVANCE TO FIELD

Narrowing the Achievement Gap in Reading and Science Learning

Recent National Assessment of Education Progress (NAEP) test results indicate that while the majority of American youth reach basic literacy levels, few are gaining the literacy knowledge, skills, and dispositions that would enable them to successfully engage in higher level, problem-solving literacy of the kind required in an information generating and information transforming economy (Donahue, et al., 1999; Mullis et al., 1994). Further, there is a persistent achievement gap between mainstream populations and those who are socio-economically, ethnically, culturally, or linguistically outside of that mainstream (Gee, 1999; Jencks & Phillips, 1998; Snow et al., 1998). Nationally, achievement gaps among different populations of students in reading are echoed by similar gaps in science learning and achievement (Donohue, et al, 1999). Low-performing students are often placed into skills-based classrooms with the well-intentioned goal of increasing their literacy proficiency. However, documentation of these differential learning opportunities for poor and minority students suggests that isolated skills-based instruction in reading may perpetuate low literacy achievement rather than accelerate literacy growth (e.g. Allington & McGill-Franzen, 1989; Haycock, 2000; Hiebert, 1991; Hull & Rose, 1989; Knapp & Turnbull, 1991).

In the current policy environment, low-performing schools are under state and federal mandates to provide literacy interventions for the lowest scoring youth in the system. Withdrawing adolescents from instruction in science to remediate reading difficulties threatens to further exacerbate historic inequities in achievement for populations of students traditionally underrepresented in the science (Barton, 2003). The need is urgent to investigate how the integration of research-based reading instruction into science learning at the high school level effects the reading and science achievement of underperforming youth.

While a good deal of recent research points to important elements of successful work in the area of adolescent literacy development (e.g. Rycik & Irvin, 2001), there are very few studies of subject area classrooms, particularly high school science classrooms, in which these effective features are being enacted (cf. Calfee & Miller, 2004; Greenleaf, Brown & Litman, 2004; Moje, et al., 2004). Even more rare is an attempt to link increases in literacy achievement to subject-area learning, particularly at the high school level. In the proposed study, we aim to investigate the relationship between growth in reading achievement and biology learning, in the context of high school biology classrooms integrating reading instruction into science teaching.

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 sorting mechanisms that track students into college-bound and non-college bound courses of study at the high school level (Hull & Rose, 1989; Knapp, 1995; Oakes, 1985; Sizer, 1992). As early as 3rd and 4th 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, et al., 1991). Students' learning outcomes in the subject areas are often measured through standardized achievement tests that require specific subject-area knowledge and 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

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 all texts demands the use of these conventions to navigate layers of meaning (e.g. New London Group, 1996; Scott, 1993; Scribner & Cole, 1981). Literacy practices become increasingly specialized throughout the school career, reflecting the broader activities that characterize the academic disciplines (Lemke, 1990; Wineburg, 1991). Skillful reading of science texts mirrors the kinds of thinking characteristic of science exploration and reasoning (e.g. Baker, 1991; Borasi & Seigel, 2000; Greenleaf, Brown, & Litman, 2004; Hynd, 1998; Lemke, 1996; Moje, et al., 2004; Roth, 1991; Their & Davis, 2002). The implications of this view for the literacy 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 (e.g., Greenleaf, Hull, & Reilly, 1994; Lee, 1995; Moje, Dillon, & O'Brien, 2000).

However, skillful reading at early grade levels will not automatically translate into higher-level academic literacy (Greenleaf, et al., 2001; Heller & Greenleaf, 2007; Lee & Spratley, 2009; Snow, 2002). Literacy researchers have 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 (e.g. Delpit, 1995; Fielding & Pearson, 1994; Freedman, Flower, Hull, & Hayes, 1995; Gee, 1999; Hillocks, 1995; Lemke, 1996; Pressley, 1998). This research has underscored the necessity of 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 and society.

Integrating Literacy Apprenticeships into Subject-Area Teaching

Some have adopted the metaphor of "cognitive apprenticeship" to describe a type of 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 (Pearson, 1996; Freedman et al., 1995). To help students develop as readers and writers, teachers create "literacy apprenticeships," engaging students in meaningful and complex literacy practices while demystifying these literacy practices (Brown et al., 1989; Lee, 1995; Osborne, 2002).

Ample studies over the past few decades have demonstrated that integrating the explicit teaching of comprehension, text structures, and word-level strategies into compelling sense-making activities with texts increases student reading achievement (Baumann & Duffy, 1997; Beck, McKeown, Hamilton, & Kucan, 1997; Guthrie, McGough, Bennett, & Rice, 1996; 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 reading and writing science content 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). Similarly, the report of the National Reading Panel (2001) concluded that teaching a combination of reading comprehension techniques is the most effective method to increase reading comprehension, rather than teaching individual comprehension strategies in isolation from one another and from content instruction (2001, p. 4 – 52). Student collaboration has also been identified as key to improving literacy achievement in the report of the National Reading Panel (2001) as well as the RAND (Snow, 2002) report.

An Intervention for Integrating Science and Reading Instruction

Based on this research in learning and reading, to support teachers' learning and adolescents' discipline-specific literacy development, Greenleaf and Schoenbach and their colleagues have developed, implemented, and studied the impact of an instructional model for academic reading instruction – the Reading Apprenticeship instructional framework (Schoenbach, et al., 1999; Greenleaf, et al., 2001). Based on work with science teacher leaders, this instructional framework has been adapted to be consistent with the specific instructional goals of secondary science. In this instructional model, the social, personal, cognitive, and knowledge-building dimensions of classroom life are woven into science teaching over time through increased reading opportunities and collaborative, metacognitive routines.

Reading Apprenticeship is a coherent instructional model that draws together research-based practices in reading instruction, including methods of engaging students in extensive reading; integrating explicit teaching of comprehension strategies; establishing relevance and making personal connections to reading materials and curriculum activities; identifying and using a variety of text structures to support comprehension; and supporting collaborative sense-making activities with written materials. Reading Apprenticeship draws both on what teachers know and do as readers in particular domains of science, and on adolescents' underestimated strengths as learners. Teachers improve students' general reading comprehension and understanding of science curriculum materials by: engaging students in more science reading; making the teacher's science-specific reading processes and knowledge visible to students; making students' reading processes, knowledge, understandings and (mis) conceptions visible to the teacher and to one another; helping students gain insight into their own reading processes as a means of gaining strategic control over these processes; and helping students acquire a repertoire of science-specific problem-solving strategies for deepening comprehension of curriculum materials. The framework centers on metacognitive conversation, involving explicit metacognitive routines, modeling, small group work, and class discussions that focus on how to read science and why people read science materials in the ways they do, as well as the science content of what is read in science classes.

Studies of the impact of Reading Apprenticeship have demonstrated increased reading achievement and academic engagement across a diverse group of adolescents enrolled in an Academic Literacy course in ninth grade (Corrin, et al., 2009; Greenleaf, et al., 2001; Schoenbach, Greenleaf, Cziko, & Hurwitz, 1999). These results have been replicated in additional studies, demonstrating that teachers' implementation of Reading Apprenticeship results in significant gains for students across varied grade levels and subject areas (www.wested.org/stratlit). 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, Brown, & Litman, 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.

A long history of research in reading has demonstrated that reading comprehension strategies are not often taught in content 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, 1984; Fielding & Pearson, 1994; Richardson, 1994; Snow, 2002). Therefore, to assure that reading instruction becomes tightly integrated into biology teaching, professional development must demonstrate features of high quality learning for teachers that are known to be effective in producing changes in classroom instruction (Strickland & Kamil, 2004). These include a focus on instruction, opportunities for continued reflection on practice, analysis of student learning on valued tasks, sustained learning opportunities for teachers, and frequent opportunities to approach pedagogical and conceptual tasks from the point of view of learners in order to build pedagogical content knowledge (Ball & Cohen, 1999; Little, 2001; Guskey & Huberman, 1996).

The Professional Development Intervention

Professional development in implementing Reading Apprenticeship is the intervention for this study of teacher practice and student learning in experimental and control classrooms. In previous studies, this instructional model, as well as the professional development methodology employed, has been effective in changing teachers' knowledge and classroom practice increasing students' literacy achievement (see Greenleaf & Schoenbach, 2004). Studies of the model have demonstrated 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;

view content-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).

To support changes in belief and practices for teachers, Reading Apprenticeship professional development is highly designed and mediated through skillful facilitation with carefully prepared text and science activities. Professional development institutes are designed to build teachers' understanding of the Reading Apprenticeship framework for advancing student literacy, develop an experiential understanding of reading processes and increase knowledge and ability to integrate supports for literacy into their own biology instruction. From the outset, teachers are immersed in a constructivist learning experience of science literacy by engaging in rich investigations into science reading and science investigation. The Reading Apprenticeship instructional framework positions literacy as inquiry, and professional development activities aim to draw on the similarities of science inquiry processes and literacy. To increase teachers' capacities to 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.

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 capabilities of diverse students, and little appreciation of the role of reading and science texts in science learning. To change the nature of instruction, it is essential to change the theories that inform teachers' decision-making processes, so that instruction becomes based less on assumptions and more on knowledge of the field and observations of real conditions in the classroom. Teachers need professional knowledge that is generative; that is, knowledge that will allow them to invent, or generate, useful responses to students' thinking in the moment (Schoenbach & Greenleaf, 2008). In the classroom, teachers respond to student thinking from their current understanding of the range of resources and needs that readers bring to text, the demands of a particular text, their pedagogical content knowledge of the topic at hand, the social interactions in the classroom, combined with their knowledge of the kinds of strategic thinking and problem solving resources that are useful in a given situation. The primary goal of Reading Apprenticeship professional development is to help teachers build a new set of analytical and thinking tools such that teachers can, and will, teach literacy in science, responding productively to their students based on their literacy needs as they read and learn in the science classroom. In order to build teacher capacity for this kind of responsive teaching, professional development includes three types of inquiries designed to build teachers' knowledge of reading, insight into student thinking and generative ability to provide effective strategy instruction.

Inquiry Designs for Building Generative Knowledge of Reading

To build teachers' knowledge of reading that will support their ongoing learning about reading as they teach, Reading Apprenticeship professional development engages teachers in a number of inquiries designed to surface and discuss thinking processes that teachers use while reading, in order to develop teacher capacity to participate in metacognitive conversations. These instructional 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 (and later students in classrooms) 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. Professional development facilitators help teachers to label their reading processes, developing 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 builds 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 will support teachers to sustain an ongoing inquiry into reading processes in their classrooms. These inquiries support their ongoing learning about reading processes as they help their students build and use new strategic approaches to comprehension challenges, all while reading science.

Inquiry Designs for Building Insight into Students' Learning

To build insight into student learning, Reading Apprenticeship professional development engages teachers in observing students reading in video and written case studies. 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 reading and student thinking, during a case discussion individual teachers will give voice to conflicting conclusions about a student. These discussions are carefully facilitated to drive teachers into a practice we intend for them to master: using classroom conversations about reading as data (formative assessment) and making evidence-based claims about students' reading and learning strengths and needs. Interpreting student thinking based on observations during reading activity helps teachers develop insights into student thinking during the case, and later develop new insights into their own students as they listen to students discuss science readings and look at their reading-related work.

Inquiry Designs for Building Effective Use of Strategy Instruction

In Reading Apprenticeship professional development sessions, the choice of texts and reading tasks 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 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 to teachers in order to build knowledge of how to support metacognitive conversation about reading processes in the classroom. Teachers are asked to describe the instructional supports for metacognitive conversation that were present in the activity: supports for reading, for thinking about reading, and for 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 a purpose for learning how to implement these instructional routines in their own classrooms. In this way, teachers solve problems of practice in collaboration with their colleagues as they design instruction that supports literacy in science.

Between sessions, teachers are asked to use these new pedagogical tools in their biology classes and practice new ways of responding to students' learning based on more informed knowledge of reading and learning. In future sessions, they bring back and analyze students' responses to their instruction then plan instructional responses to student work in collaboration with their colleagues.

New Knowledge, New Practices

Through this cycle of inquiry, we aim 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" (NSDC, p. 29). We hypothesize that teachers develop new knowledge and resources about text, science reading and student thinking that supports their teaching of reading in science. With practice, teachers learn to deploy these resources more flexibly, on demand, as students need them. Students then practice these thinking tools as authentic and relevant responses to real reading situations, 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 this practice, both teachers and students can stretch beyond their current ability and gain more expertise and capacity.

Logic Model

Driving the Reading Apprenticeship professional development design, then, are some key assumptions about teaching, learning and teacher change that informed the design of this study. The logic model linking professional development to classroom change is as follows:

1. By immersing teachers in a model of learning that integrates science inquiry with literacy inquiry—a model we want them to replicate adaptively in their classrooms—and by helping them to analyze and reflect on the scaffolds and supports for learning within this model, we hypothesize that teachers will build the pedagogical tools they need to create similar learning environments.
2. As teachers reflect on their own and others' learning during iterative inquiries into science reading, they will develop new concepts about reading, about readers and about teaching and learning; their changing understandings will become more reflective of the sociocultural and constructivist approach to science literacy that underlies the Reading Apprenticeship model.
3. These new concepts will lead to teachers' developing new attitudes and dispositions toward the social context of their classrooms, toward science texts and toward the capacities of students as learners, holding more thoughtful and informed views.
4. These new pedagogical tools, concepts, attitudes and dispositions will give teachers both the skill and the will to create new learning opportunities for students to read more science texts, read with more support and read more scientifically in their science classrooms, opportunities that will be different than those available to students of teachers who have not had Reading Apprenticeship training.

RESEARCH QUESTIONS

Prior studies of professional development have suffered from a lack of scientific tools to demonstrate impacts on teacher knowledge and classroom practices, and to link these impacts to student engagement and achievement. In this study, a group randomized, experimental design was used to assess program impacts on high school biology teaching and learning in schools serving traditionally low-achieving students. The study relied on a set of innovative pre and post-intervention assessments of teacher knowledge, teacher practices in science and literacy, and teacher instructional beliefs; as well as post-intervention assessments of student learning in biology and reading comprehension.

We investigated the effects of integrating literacy instruction with biology coursework on student learning in both subjects, targeting students historically underrepresented in the sciences. A randomized controlled study was designed to test the following hypotheses:

> H1: 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.
>
> H2: Students in experimental classrooms will demonstrate greater increases in science understanding, reading proficiency, and engagement in science learning than their counterparts in control classrooms.

In testing these hypotheses, the study methods included multiple measures of both the students' opportunity to learn and student learning. A qualitative study of a smaller sample of teachers, including classroom observations and interviews, was used to validate and explain quantitative findings and to identify factors that influence the success of the pedagogical approach. The study also enabled us to investigate the extent to which these instructional methods have different impacts for groups of students historically underrepresented in the sciences.

DESCRIPTION OF INTERVENTION

Experimental Condition

Teachers randomly selected to be in the experimental condition received professional development in Reading Apprenticeship and support to integrate science content and reading instruction. The 10-day training session, with certified Reading Apprenticeship professional development providers, was designed to utilize the inquiry tools and approaches to professional development in content area reading developed by the Strategic Literacy Initiative (Greenleaf & Schoenbach, 2004). In the summer of 2005, teachers participated in five days of training. 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 email 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 three-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 list serve, moderated by the professional development coaches. See **Appendix A** for an overview of activities carried out in the 10 days of professional development.

To support implementation and to assure equal access across experimental sites to opportunities to read in science, these 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 non-fiction selections linked to the biology topics and state curriculum frameworks. For a sample of state-recommended reading materials in science, see www.cde.ca.gov.

Stipends covered teacher participation in the professional development, including travel, food, and housing for the summer institute, honoraria for teacher's time, additional stipends or substitutes for day-long sessions during the school year, and up to $200 for instructional reading materials.

Control Condition

Teachers 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. Participation in professional development activities, changes in teaching practices, acquisition of knowledge and skills, and other changes in conditions and circumstances were tracked and monitored in control group sites.

EXPERIMENTAL DESIGN, RECRUITMENT, AND DATA COLLECTED

The principal aim of the study was to test the effectiveness of teacher training in the integration of reading instruction and science content on teacher knowledge and skills, instructional practices, and on student achievement in science and reading. A *true, group-randomized, experimental design* was designed to control for most threats to internal validity (Cook & Campbell 1979, Murray 1998). Schools and the participating teachers within them were randomly assigned to one of two different groups – an experimental group and a wait-listed control group – with a minimum of 25 schools per group as shown in **Table 1**. Standard regression- and mixed-modeling procedures were used to detect treatment effects on teacher practices and student achievement.

Table 1. Experimental Design

	Year 1 **2005/06** Sum			Spring	*Year 2* **2006/07** Sum		Spring
Teachers							
9th/10th Biology							
Group #1	O			O			O
Group #2	O	*PD*	*RA*	O	*PD*	*RA*	O
Students							
8th Grade							
Group #1				O			
Group #2				O			
9th Grade							
Group #1				O			O
Group #2				O		*RA*	O
10th Grade							
Group #1							O
Group #2						*RA*	O

O = observations or measurement points
PD = Reading Apprenticeship Professional Development
RA = Classroom implementation of RA
TxU = Treatment as Usual Condition

Recruitment of Schools and Teachers

The target population was high school biology teachers and their students in public high schools across California. To contribute to the STEM goals of NSF, the study took place in high schools in California that serve populations of students historically underrepresented in the advanced sciences. The sample consisted of schools with high proportions of these students to better ascertain the impact of integration of literacy instruction with biology course-work for groups of students historically underrepresented in the sciences. Schools, not teachers, served as the unit of randomization to minimize contamination of the control group through teacher interaction. Prior to randomization, participating high schools were pair-matched with similar schools based on 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).

Schools were randomly assigned to treatment and control groups within each pair of schools. The SCI is based on the following factors: student mobility (percent of students who first attended school in current academic year), ethnicity (percent of students in seven ethnic/race categories), average parental education, percent receiving subsidized meals, percent of teachers fully credentialed, percent of teachers with emergency credentials, percent of English language learners, average class size, and year-round school status. In creating the index, each factor was weighted proportional its relationship to the California's Academic Performance Index, based on a linear regression model. To control for the effect of experience, all teachers were credentialed in biology and had taught for at least 3 years at the initiation of the data collection phase of the study.

Roles of Research Team

Because one of the PI's 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 the 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 vs. professional development responsibilities. The primary role of UCLA's CRESST Center was to develop, field test, and analyze measures of the nature and degree of literacy instruction in biology classrooms, including teacher surveys and teacher assignments. They also worked with the Strategic Literacy Initiative research team to develop and pilot a performance assessment of student biology and literacy learning. All scoring of these measures was carried out by CRESST. 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 which only they had access to. All statistical analyses of data were done by either CRESST or the Los Alamitos Research and Evaluation office of WestEd.

Data Collection

Several types of data were collected to answer the research questions. These data sources included measures of student achievement and engagement, teacher surveys, analysis of teacher assignments, and observations of classroom practice. Below we describe these instruments.

Teacher Change/Implementation

Teacher Survey: Both the experimental group and the control group took a survey about their classroom practices in summer 2007, which serves as a post-test to two identical surveys taken in Summer 2005 and Summer 2006. The survey assessed six global constructs related to effective integration of literacy and biology instruction: (1) science reading opportunities, (2) collaboration, (3) metacognitive inquiry, (4) comprehension strategies instruction, (5) a feature of instruction called "negotiating success"—a focus on designing and modifying instruction and assessment to support student learning, and (6) teacher beliefs about reading, learning and diversity. The six constructs were divided into 14 sub-constructs. Sub-constructs reflected the apprenticeship model and the range of science reading opportunities offered to students, and teacher modeling, guidance and support for key reading and discourse routines, tools, strategies and dispositions. (See **Appendix B** for a summary of the survey items, constructs, and consistency studies conducted.)

Teacher Assignments: 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 teacher also submitted an in-depth coversheet for each assignment, in which they describe 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.

The use of teacher classroom assignment ratings as an indicator of practice is a methodology developed by CRESST researchers, as part of work with LAUSD (Aschbacher, 1999; Clare, 2000). CRESST research supports the validity and reliability of using these ratings as an indicator of classroom practice quality (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. Ratings involve elicitation of teacher thinking about a specific lesson.

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 six new *literacy* dimensions that measure literacy instruction, opportunities for engagement with rich text, metacognitive inquiry into reading and thinking processes and teacher support for the cognitive and metacognitive demands of the literacy task. For each rubric a 4-point scale (1 = poor, 4 = excellent) was used to rate the quality for each assignment for separate dimensions. (See **Appendix C** for a summary of the scoring process and reliability of the scoring.)

Teacher Interviews: Interviews were conducted using semi-structured protocols aligned with the teacher survey content/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. Interviews were recorded and subsequently rated on a 4 point rubric on five dimensions: reading opportunities, support for student reading engagement, metacognitive inquiry, reading comprehension routines, and collaboration (See **Appendix D** for a description of the protocol used to conduct, code, and score the teacher interview data).

Classroom Observations: Classroom observations were conducted on a small subset of 9 classrooms participating in the study. We focused our limited time and resources on observing treatment classrooms where teachers were making an effort to integrate literacy instruction with biology coursework. The purpose of the classroom observations was to provide a snapshot of student literacy learning opportunities in classrooms where teachers received professional development and support to integrate reading instruction and science content. The observations permitted us to learn about implementation in these classrooms independently of teacher self-report.

For these observations, we adapted the Classroom Observation and Analytic Protocol developed by Horizon Research, Inc. (HRI) for *Looking Inside the Classroom: A Study of K-12 Mathematics and Science Education in the United States* (May, 2003). The HRI instrument contains four components that assess the quality of the design and implementation of mathematics and science lessons: the lesson design, lesson implementation, mathematics/science content addressed and classroom culture. For this study, we collapsed elements of lesson design and lesson implementation into a single component, instructional strategies, and added a lesson component focused on literacy opportunities comprising 19 items. Key indicators within the four resulting components — instructional strategies, science content, literacy opportunities, and classroom culture — were modified from the original HRI instrument to better reflect the goals of the current study and the Reading Apprenticeship framework.

Classroom observation data were analyzed in three parts. First, we analyzed findings related to the ratings of lesson quality. Second, we compared lesson quality ratings from the nine classrooms we observed with lesson quality ratings from the national sample of classrooms participating in the Inside the Classroom study. Finally, we conducted an in depth analysis of lesson descriptions and materials designed to identify factors that influenced the success of the Reading Apprenticeship intervention. A detailed report of these analyses has been submitted as a separate document.

Student Learning Opportunities and Outcomes

Positive consent and student outcome measures, including Opportunity to Learn Surveys and state standardized test scores for the baseline and intervention year, were collected for one 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. Non-identifiable student standardized test score data was collected for each participating teacher to broaden the sample and its representativeness.

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 more standardized tests that would have been more sensitive to our intervention, such as the Performance Assessment in Science for Students 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 non-mandated test to students since so much instructional time was already devoted to testing in the state. We therefore used the California Standards Test (CST) in Biology to assess biology understanding and the CST English language arts 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 scores were collected from participating school districts beginning in the Fall of 2007. However, 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 for 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 for whom we did not obtain parental consent. For linked students we collected: CST English Language Arts and ELA Reading Comprehension for 2004/05, 2005/06, and 2006/07. The 2005/06 CST measures were used as covariates in the longitudinal impact analysis models. For students who are unlinked in the dataset, we collected CST English Language Arts, ELA Reading Comprehension, and Biology Scores for students in participating teachers classrooms in 2004/05, 2005/06, and 2006/07. The 2004/05 anonymous student test score data were aggregated to the teacher level and used as covariates in our cross-sectional student test score impact analysis models.

The English Language Arts and Reading 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.

Student Opportunity To Learn Survey: Based on prior surveys developed at WestEd for the Performance Assessments in Science (PASS) assessments, student 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 Opportunity to Learn (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' perception of themselves as readers and learners. Six key constructs were assessed by the survey and used as outcomes: (1) class emphasis on reading in biology, (2) integration of biology and literacy activity, (3) identifying as a reader, (4) student identity, (5) motivation in class, and (6) course consequences on reading science. **Appendix E** provides a map of individual items to each of the constructs and a summary of the technical quality of the instrument. The survey was administered to students in both treatment and control groups in spring of 2006/07, the intervention year.

Integrated Learning Assessment: In tracing the effects of the professional development on student reading and learning, we were eager to capture the complexity of student outcomes targeted by the intervention—increased engagement and use of metacognitive and comprehension supporting routines that support students to become self-monitoring and self-governing readers of science, as well as increased achievement in biology and literacy. We obtained supplementary funding from NSF to develop an end-of-year Integrated Learning Assessment (ILA) that affords a closer look at aspects of students' literacy and science learning related to the treatment and where we would expect to see differences between treatment and control groups.

We administered a genetics version of the assessment to students in the spring of 2007 (the treatment year) in our treatment and control classrooms. The ILA is a performance-based assessment that integrates CRESST's previous work on model-based assessment and explanation tasks with an existing Reading Apprenticeship curriculum-embedded assessment, the CERA. Students read a complex, grade-level science text, describe their reading and thinking processes, answer a set of comprehension questions and do some writing on the topic. Scoring attends both to students' comprehension and conceptual understanding, and students' literacy, scientific thinking and discourse processes.

The ILA was distributed in April 2007. Unfortunately because this instrument was funded as a late supplement to the original grant, it took until this time to design, pilot and launch the instrument with the teachers in this study. Because of the late launch, this instrument was offered as an optional piece of the data collection, coupled with the offer of an additional stipend payment. This capitalized on the good will of participating teachers, while also aiming to avoid punishing them if they could not find two days to administer this assessment during the testing season. However, the rate of return was not as high as with the other items we collected. **Appendix F** describes the process of scoring and ensuring the reliability of scores on this instrument.

Table 2 below shows the data collection schedule for the various data collection to gauge evidence of teacher implementation and student learning outcomes.

Table 2. Data Collection Schedule

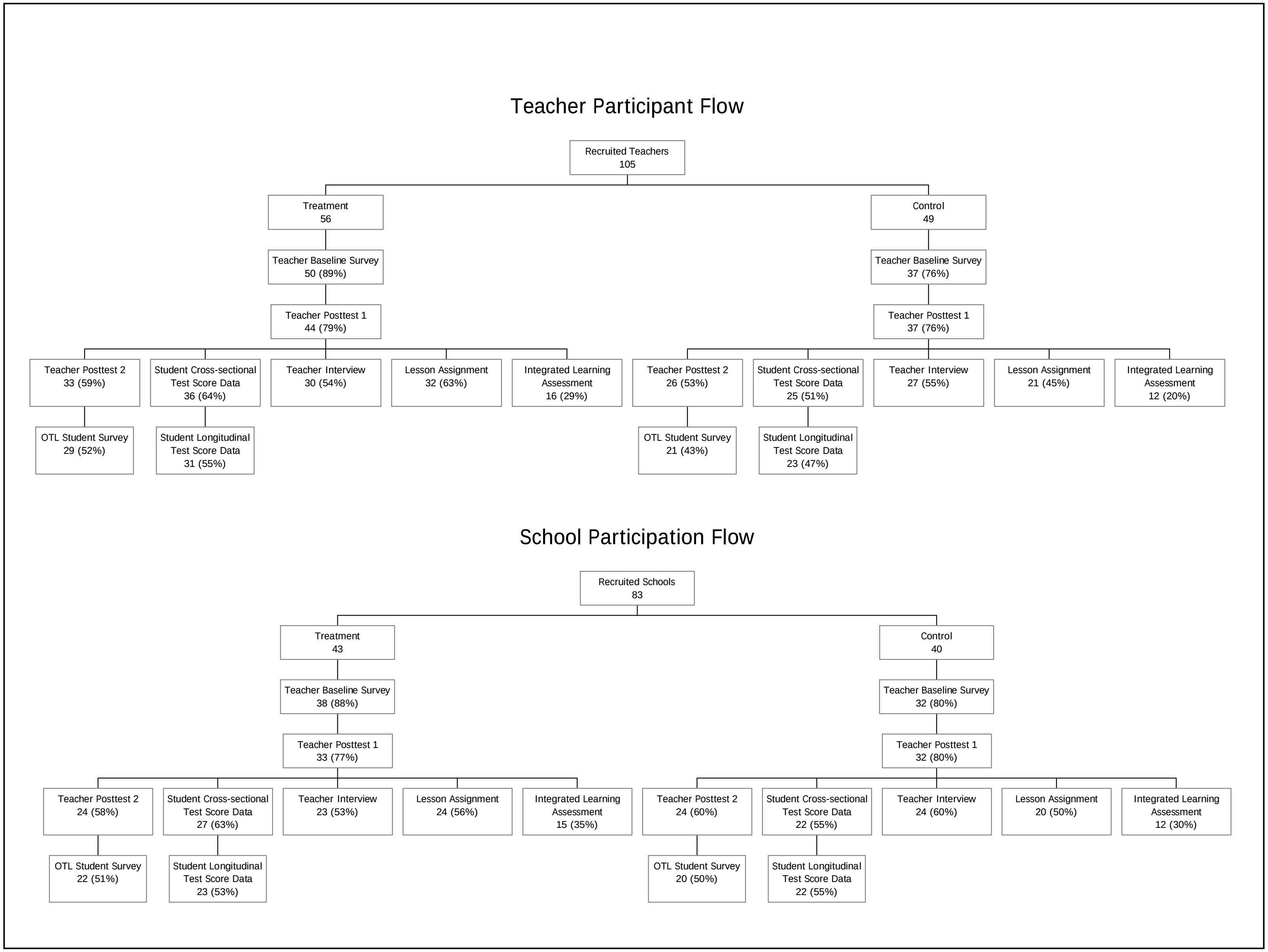
	Baseline **2004/05**	*Year 1* **2005/06**	*Year 2* **2006/07**
Teacher Practice			
Teacher Surveys (Instructional Beliefs/Practice)	Summer	Summer	Summer
Teacher Assignments (Genetics, Cell Biology)			Fall/Spring
Teacher Interviews			Spring
Student Opportunity to Learn Surveys			Spring
Student Outcome Measures			
Integrated Learning Assess. (Genetics, Cell Biology)			Spring
State Test Scores (Biology & ELA)	Spring	Spring	Spring

Retention of Schools and Teachers

Figure 1 below shows the number of teachers and schools randomly assigned to treatment and control groups, as well as the data retention rates for each data source. Overall, 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. Note that teachers and schools were recruited and randomized to condition in the spring of 2005, two to three months prior to the scheduled summer professional development institute. Schools and teachers were randomly assigned in batches so that adequate notice could be given to teachers to schedule participation in the summer professional development.

As shown in **Figure 1**, 89 percent of treatment teachers and 76 percent of control teachers provided responses on the baseline teacher survey, 79 and 76 percent of treatment and control teachers participated in the 1st-year post-implementation teacher survey, and 59 and 53 percent 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 percent of randomly assigned teachers, teacher interviews were conducted with 55 percent of teachers, and lesson assignment data were collected from 63 percent of treatment teachers and 45 percent of control teachers. Cross-sectional student test score data were collected from 64 percent of treatment teachers and 51 percent of control teachers. As discussed above, Integrated Learning Assessment (ILA) data were collected as an option from volunteer teachers. Approximately 29 and 20 percent of treatment and control teachers, respectively, participated in ILA data collection. The school participation chart at the bottom of **Figure 1** shows similar data return rates as that for teachers.

Figure 1. *Teacher and School Retention by Data Source*



Equivalence of Treatment and Control Groups

Although data attrition levels were fairly high, attrition patterns were fairly 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/control group equivalence (or lack thereof) at the time of random assignment and at subsequent data collection periods, we present school-, teacher-, and student characteristics by data source. **Table 3a** shows school characteristics by treatment/control status for the randomized sample, school and teacher characteristics for the teacher pretest sample, and school and teacher characteristics for the student OTL survey sample. Overall, the randomized and teacher pretest samples show a high degree of similarity, with few meaningful differences in school performance and demographic characteristics. The student OTL sample, which is comprised of about 50 percent of randomized schools/teachers, exhibits more evidence of treatment/control group non-equivalence than the teacher pretest sample, but none of the differences are statistically significant. Treatment schools had about 30 percent 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 3b shows pre-intervention characteristics of students in treatment and control schools based on the 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 3b** could be due to differences in teacher participation rates, student participation rates, or both factors. Group differences in pre-intervention 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 in the student level characteristics in **Table 3b** was higher than that reflected by the school aggregate characteristics in **Table 3a** (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.

Table 3a. Pre-intervention Characteristics by Treatment/Control Status for Randomized Sample, Teacher Pretest Sample, and Student OTL Sample

	Treatment	Control	Difference	p-value	Diff/SD
Randomized Sample (83 Schools)					
School Characteristics (83 Schools)					
Academic Performance Index	653.39	649.46	3.93	0.82	0.05
API State Rank	4.93	4.89	0.03	0.96	0.01
Science CST (NCE)	46.79	46.67	0.12	0.94	0.02
Free/Reduced-price Meals	37.20	37.97	-0.78	0.87	-0.04
African American (%)	11.27	9.24	2.02	0.35	0.21
Hispanic (%)	39.59	44.05	-4.46	0.42	-0.18
Asian (%)	10.13	8.90	1.23	0.68	0.09
White-NonHispanic (%)	31.62	32.13	-0.51	0.93	-0.02
English Learners (%)	19.02	17.29	1.73	0.57	0.13
Teacher Pretest Sample					
School Characteristics (70 Schools)					
Academic Performance Index	648.21	655.00	-6.79	0.71	-0.09
API State Rank	4.72	5.17	-0.44	0.52	-0.16
Science CST (NCE)	46.20	47.19	-0.99	0.54	-0.15
Free/Reduced-price Meals	38.03	35.80	2.23	0.68	0.10
African American (%)	11.25	9.40	1.85	0.44	0.19
Hispanic (%)	41.27	42.35	-1.07	0.86	-0.04
Asian (%)	9.76	9.91	-0.15	0.96	-0.01
White-NonHispanic (%)	30.02	31.87	-1.86	0.76	-0.07
English Learners (%)	19.44	16.97	2.48	0.48	0.18
Teacher Characteristics (87 Teachers)					
Female	0.62	0.75	-0.13	0.20	-0.28
Biology Major	0.26	0.32	-0.07	0.59	-0.14
Years Teaching Science	9.39	8.32	1.07	0.35	0.20
Years in School	6.58	6.14	0.44	0.77	0.09
Student OTL Survey Sample					
School Characteristics (42 Schools)					
Academic Performance Index	640.36	651.76	-11.39	0.63	-0.15
API State Rank	4.50	5.05	-0.55	0.54	-0.19
Science CST (NCE)	45.54	47.25	-1.72	0.41	-0.27
Free/Reduced-price Meals	38.70	35.89	2.81	0.69	0.13
African American (%)	10.15	8.97	1.18	0.68	0.13
Hispanic (%)	46.56	42.79	3.77	0.64	0.15
Asian (%)	8.26	12.11	-3.85	0.41	-0.42
White-NonHispanic (%)	28.03	29.95	-1.92	0.79	-0.08
English Learners (%)	21.05	15.95	5.10	0.23	0.36
Teacher Characteristics (50 Teachers)					
Female	0.55	0.71	-0.16	0.25	-0.33
Biology Major	0.25	0.42	-0.17	0.24	-0.36
Years Teaching Science	9.28	7.48	1.80	0.23	0.34
Years in School	6.45	5.38	1.07	0.51	0.23

Notes: p-values are based on multilevel regression models in which treatment group status is included as a covariate. Effect sizes calculated by dividing group difference by the pooled standard deviation.

Table 3b. Pre-intervention Characteristics of Students in Treatment and Control Schools

	Treatment	Control	Difference	p-value	Diff/SD
Longitudinal Test Score Sample (45 Schools)					
Student Characteristics					
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 05 (std)	-0.08	0.11	-0.18	0.53	-0.19
Mathematics CST 05 (std)	-0.10	0.17	-0.27	0.55	-0.27
Cross-sectional Test Score Sample (51 Schools)					
Student Characteristics					
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 04	**321.72**	**329.56**	**-7.84**	**0.16**	**-0.52**
Teacher Reading Comp 04 (std)	**-0.07**	**0.04**	**-0.11**	**0.08**	**-0.44**
Student OTL Survey Sample (42 Schools)					
Student Characteristics					
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

Notes: p-values are based on multilevel regression models in which treatment group status is included as a covariate. **Bolded** numbers correspond to treatment/control differences that may be substantively meaningful. Effect sizes calculated by dividing group difference by the pooled standard deviation.

The treatment/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. Recall that because the anonymous cross-sectional data were linked to teachers, but not to specific students, we did not have pre-intervention test score data <u>for the students</u> in the cross-sectional sample. Instead, we have anonymous, pre-intervention data for <u>2004/05 students served by the teachers</u> in the analytic sample. As shown in the middle panel in **Table 3b**, prior to the intervention, the students served by treatment teachers exhibited substantially lower biology and reading comprehension test scores than those served by control teachers – with group differences of between 40 and 50 percent of a standard deviation. Overall, for both sets of test score samples, the results suggest that differential attrition may have led to treatment/control group non-equivalence. Note, however, that fewer treatment/control school differences were apparent based on the student OTL survey sample.

OUTCOME ANALYSES

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; Raudenbush & Bryk, 2002; Murray, 1998). In each of the impact analyses, we controlled for baseline (pre-test) measures of outcome variables when available, randomization strata (i.e., pairs), and student-level covariates when analyzing student outcomes. In addition to examining main effects for program impacts on student performance outcomes, we also examined differences in impacts for the following subgroups: (1) English learners and English proficient students (test scores only), (2) females and males, (3) racial/ethnic groups, and (4) and low- and high-performing students. For the outcomes assessed with student OTL surveys, we examined differential impacts across student gender, race/ethnicity, and student home language (English and non-English).

H1: Teacher Outcomes: Integration of Literacy into Instructional Practice

Teacher Surveys: Analysis of pre- and post- 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 4.**

To summarize these results, there is little evidence that the means on pre-intervention survey measures differ by treatment/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 7 of the 14 sub-constructs shown in **Table 4** at the end of the intervention year. These constructs included:

- *Science Reading: Content*, the extent to which science content from reading materials is acquired through student work and meaning making (versus delivered directly by the teacher through lecture);
- *Collaborative Activities, Teacher Modeling:* the extent to which teachers modeled and supported collaboration instructionally;
- *Collaborative Activities, Student Practice:* the extent to which students had access to one another for core work;
- *Metacognitive Inquiry, Teacher Modeling:* the extent to which teachers modeled metacognitive inquiry and reading routines;
- *Metacognitive Inquiry, Student Practice:* the extent to which students had opportunities to practice metacognitive inquiry and reading routines;
- *Comprehension Strategies, Teacher Modeling:* the extent to which teachers provided

modeling and explicit instruction in comprehension-supporting strategies; and

- *Teaching Philosophy, Reading:* the extent to which teachers believe reading plays a vital role in science learning.

Table 4. Treatment/Control Differences in Post-Surveys

	Treatment	Control	Difference	p-val	Diff/SD
Teacher Survey – 2nd Post-Survey					
Student Reading Opportunities - Texts	3.11	3.17	-0.06	0.42	-0.13
Science Reading Opportunities - Learning Structure	3.09	2.78	0.30*	0.09	0.51
Science Reading Opportunities - Content	3.19	2.85	0.34**	0.01	0.78
Collaboration - Teacher Modeling	3.30	2.82	0.49**	<0.01	0.64
Collaboration - Student Practice	3.13	2.77	0.36**	0.01	0.61
Metacognitive Inquiry - Teacher Modeling	3.15	2.62	0.52**	<0.01	0.77
Metacognitive Inquiry - Student Practice	3.07	2.27	0.80**	<0.01	1.16
Comprehension Strategies -Teacher Modeling	3.55	2.85	0.70**	<0.01	0.72
Comprehension Strategies - Student Practice	3.14	2.97	0.17	0.52	0.25
Negotiating Success - Instruction	3.61	3.46	0.15	0.15	0.29
Negotiating Success - Assessment	3.11	2.92	0.19*	0.08	0.32
Teaching Philosophy - Reading	4.27	4.05	0.23**	0.05	0.66
Teaching Philosophy - Learning	3.45	3.38	0.07	0.10	0.17
Teaching Philosophy - Diversity	3.95	4.07	-0.12	0.11	-0.28

** statistically significant at the .05 level
* statistically significant at the .10 level

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 that 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 0.61 to 1.16 standard deviation units. Moreover, intervention group teachers reported higher levels on three other constructs – *Science Reading Opportunities-Learning Structures* (p=0.085) – indicating that students accessed the content of science texts through reading and class discussion rather than through teacher lecture, *Negotiating Success-Assessment* (p=0.079) – indicating more use of assessment to drive instruction, and *Teaching Philosophy-Reading* (p=0.053) – a belief in the value of reading in biology learning, although these increased levels were not statistically significant at conventional levels (p < .10). Ancillary analyses indicated that the differences in post-survey outcomes are unlikely due to treatment/ control group differences in sample selectivity, as no differences were apparent between treatment and control group teachers on pre-test measures among the sample of teachers with non-missing 2nd year post-survey data.

Teacher Assignments: Table 5 shows means on the 4-point scoring rubric for the Cell Biology and Genetics classroom assignment ratings. For the Cell Biology assignment, treatment/control differences in ratings were not statistically significant at conventional levels, although teachers in the intervention group received higher ratings on *Cognitive Challenge* (p-value = 0.064) and *Support for Cognitive Challenge* (p-value = 0.057) in the literacy dimension. However, it is worth noting that treatment teachers' ratings were more than one-half of a standard deviation higher than the ratings of control teachers. Thus, there is some evidence that intervention group teachers offered more challenging literacy opportunities and provided more appropriate support for literacy tasks. For the Genetics assignment, teachers in the intervention group received higher ratings for *Support for Cognitive Challenge* in both the content and literacy domains. They also received higher ratings on the Literacy Goals dimension (p-value = 0.094), suggesting that their assignments provided more explicit and elaborated goals for student literacy.

Table 5. Teaching Assignment Differences by Treatment/Control Group

	Treatment	Control	Difference	p-value	Diff/SD
Cell Biology Assignment					
Content					
Goals	2.91	3.21	-0.30	0.17	-0.40
Cognitive Challenge	2.72	2.68	0.03	0.99	0.05
Support for Cognitive Challenge	2.81	3.05	-0.24	0.27	-0.32
Literacy					
Goals	2.38	2.05	0.32	0.17	0.40
Cognitive Challenge	2.47	2.16	0.31*	0.06	0.54
Support for Cognitive Challenge	2.53	2.05	0.48*	0.06	0.57
	Treatment	Control	Difference	p-value	Diff/SD
Genetics Assignment					
Content					
Goals	3.03	2.81	0.22	0.32	0.29
Cognitive Challenge	2.87	2.71	0.15	0.39	0.23
Support for Cognitive Challenge	3.17	2.81	0.36**	0.05	0.57
Literacy					
Goals	2.67	2.19	0.48*	0.09	0.53
Cognitive Challenge	2.40	2.19	0.21	0.24	0.31
Support for Cognitive Challenge	2.87	2.33	0.53**	0.02	0.70

** statistically significant at the .05 level
* statistically significant at the .10 level

Teacher Interviews: As described above, teacher interviews were recorded and subsequently rated on a 4 point rubric on 5 dimensions: reading opportunities, support for student reading engagement, metacognitive inquiry, reading comprehension routines, and collaboration. A sixth dimension, inquiry science, was developed after the interviews were conducted and was coded as a dichotomous outcome. **Table 6** below shows mean ratings for treatment and control teachers. The results indicate that teachers in the intervention group exhibit substantially higher interview ratings than the 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*. Thus, although the kinds of reading students are asked to do—the range of texts and volume of reading expected—do not differ in intervention and control classrooms at conventional levels of statistical significance (p=0.103), 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 0.95 to 1.47 standard deviation units.

Table 6. Teacher Interviews: Differences by Treatment/Control Group

	Treatment	Control	Difference	p-val	Diff/SD
Interview Outcomes					
Reading Opportunities	2.88	2.61	0.27	0.10	0.45
Support for Student Reading Engagement	2.92	2.17	0.75**	0.01	0.95
Metacognitive Inquiry	2.55	1.63	0.92**	<0.01	1.47
Reading Comprehension Routines	3.00	2.07	0.93**	<0.01	1.19
Collaboration	2.98	2.19	0.80**	<0.01	1.42
Inquiry	0.13	0.11	0.02	0.72	0.07

** statistically significant at the .01 level

In summary, several sources of data 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.

Teacher Implementation: As described above, Reading Apprenticeship is a multifaceted instructional framework designed to influence instructional practices across a broad array of areas. Because the framework is comprehensive, explicitly describing the level of teacher implementation of Research Apprenticeship-aligned practices is a challenge. We relied on the teacher interview data to describe the level of program-aligned practices among teachers in the treatment and control groups. We used the teacher interview data as the basis for our implementation measure because it was designed to assess each critical domain consistent with the Reading Apprenticeship model – (a) science reading opportunities; (b) teacher supported in-class reading; (c) metacognitive inquiry; (d) specific comprehension routines, strategies, and processes; and (e) student-student collaboration. Moreover, each of these qualities was assessed with a scoring metric that allowed the development team to authentically classify the level of implementation (see **Appendix D** for a description of the interview scoring rubric).

Table 7 describes the level of implementation among all teachers interviewed and separately for treatment and control teachers, based on the ratings of teacher interviews.[1] The results for the control group in the table represent estimates of usual practices of teachers who have not been trained in Reading Apprenticeship. Overall, approximately two-thirds of control group teachers exhibited *low levels* of implementation in the areas of teacher support for student reading, comprehension strategies, and collaboration. The least frequently implemented area was metacognitive inquiry, where nearly 90 percent showed *low* implementation. The most frequent literacy practice employed by control group teachers was providing students with science reading opportunities – where 64 percent exhibited *medium levels* of implementation and 12 percent (3 teachers) engaged in *high levels* of implementation.

Table 7. Level of Implementation as Assessed by Teacher Interviews

	All		*Treatment*		*Control*	
	Count	%	Count	%	Count	%
Overall Implementation						
Low	22	37.9	6	18.2	16	64.0*
Medium	24	41.4	16	48.5	8	32.0
High	12	20.7	11	33.3	1	4.0
Science Reading Opportunities						
Low	13	22.4	7	21.2	6	24.0
Medium	34	58.6	18	54.6	16	64.0
High	11	19.0	8	24.2	3	12.0
Teacher Support for Student Reading						
Low	24	41.4	7	21.2	17	68.0*
Medium	19	32.8	12	36.4	7	28.0
High	15	25.9	14	42.4	1	4.0
Metacognitive Inquiry						
Low	32	55.2	10	30.3	22	88.0*
Medium	22	37.9	19	57.6	3	12.0
High	4	6.9	4	12.1	0	0.0
Comprehension Strategies						
Low	24	41.4	7	21.2	17	68.0*
Medium	18	31.0	12	36.4	6	24.0
High	16	27.6	14	42.4	2	8.0
Collaboration						
Low	19	32.8	3	9.1	16	64.0*
Medium	27	46.5	18	54.5	9	36.0
High	12	20.7	12	36.4	0	0.0

*Treatment/control group difference is statistically significant at the .05 level.

[1] Scores on each of these ratings were classified to reflect low, medium, and high levels of implementation.

As shown in **Table 7**, Treatment teachers exhibited significantly higher levels of implementation than control teachers in all areas except science reading opportunities, which, again, was the most frequently implemented practice employed by control teachers. About 40 percent of intervention teachers exhibited *high levels* of implementation in the areas of teacher support for student reading, comprehension strategies, and collaboration. As was the case with control teachers, practices involving metacognitive inquiry were the least common among treatment teachers. Still, 58 percent and 12 percent of treatment teachers exhibited medium and high implementation levels, respectively, in this area – substantially higher levels than that of control teachers. Overall, the results in **Table 7** suggest that the intervention was consistently associated with increases in Reading Apprenticeship-aligned practices in all areas except science reading opportunities in which they were similar to the control group.

This study targeted students who are least well represented in the sciences and in institutions of higher education: students from lower SES backgrounds, English learners, and African-American and Latino students. Therefore, we examined the extent to which the level of implementation was related to the types of students served in treatment and control schools, by tabulating school performance and school-level demographic characteristics by the level of teacher implementation. As shown in **Table 8**, treatment teachers with low implementation were more likely to teach in schools that exhibited lower levels of school performance; and served higher proportions of Latino students, students eligible for free/reduced-price meals, English learners, and students whose parents had lower levels of education. Low implementers were also in more likely to be in schools that served lower proportions of white students than medium- or high implementers. Conversely, treatment teachers who implemented Research Apprenticeship-aligned practices at a high level were more likely to be in schools serving students who traditionally perform well in school than their counterparts who implemented at low levels. These results suggest that high implementation is more likely to take place in school environments with fewer barriers to academic performance.

Table 8. School Demographic Characteristics by Level of Implementation

	Treatment Schools			Control Schools		
	Low	*Med*	*High*	*Low*	*Med*	*High*
School Characteristics						
API (2008)	611.7	695.6	707.6*	735.1	704.0	495.0
African American (%)	13.3	7.9	6.8	9.7	4.9	25.7
Asian (%)	15.3	4.9	8.9	15.4	4.9	7.3
Latino/a (%)	60.5	47.1	45.1	39.9	51.3	61.2
White (%)	8.8	31.2	32.0*	28.7	34.4	1.7
Free/reduced-price meals (%)	73.7	45.1	40.3**	36.4	42.3	77.7
English Learners (%)	35.2	19.9	18.8*	12.9	14.3	34.7
Parental Education	2.0	2.7	2.8**	2.9	2.7	2.1
Number of schools (teachers)[A]	4(6)	9(16)	10(11)	12(16)	4(8)	1(1)

*Implementation group difference is statistically significant at the .10 level.
**Implementation group difference is statistically significant at the .05 level.
[A]Note that more than one teacher can be present in a particular school, i.e., schools may be double counted across columns in the table.

Note, however, that there is no discernable relationship between the level of literacy implementation in biology classes and school characteristics in the control schools, although the limited distribution of implementation levels in the control group (i.e., the small numbers of teachers with medium- and high levels of implementation) may be masking a relationship. Overall, the distribution of implementation levels among teachers in the treatment group suggests that teachers find it easier to implement these new practices when serving a more advantaged group of students.

H2: Student Outcomes: Science Understanding, Reading Proficiency, and Engagement in Science Learning

Student Opportunity To Learn Survey: To investigate treatment/control group differences on the OTL survey outcomes, we estimated multi-level 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). To some extent, the student OTL survey results presented in **Table 9** 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 & Literacy*—a measure of student practice of comprehension supporting routines and strategies. Moreover, students in treatment schools reported higher levels on the *Student Identity* construct than students in control schools (p=0.054).

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. 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 non-English speaking families than for those from English-speaking families – with four out of the six OTL outcomes being statistically significant for non-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 non-English background students range from 0.34 to 0.43 standard deviations.

Table 9. Student Opportunity to Learn Surveys by Treatment and Control Group

	Treatment	Control	Difference	p-value	Diff/SD
OTL Outcomes					
Reading in Biology	3.01	2.85	0.16**	0.05	0.30
Integration of Biology & Literacy	2.80	2.61	0.19**	0.01	0.30
Identifying as Reader	2.56	2.54	0.02	0.81	0.03
Student Identity	2.84	2.72	0.12*	0.05	0.18
Motivation in class	2.88	2.80	0.09	0.16	0.14
Reading Science	2.79	2.70	0.09	0.33	0.11
OTL Outcomes by Home Language					
Reading in Biology					
Non-English	3.01	2.80	0.21**	0.03	0.40
English	3.01	2.87	0.14	0.11	0.25
Integration of Biology & Literacy					
Non-English	2.82	2.55	0.27**	0.00	0.43
English	2.78	2.63	0.15**	0.04	0.24
Identifying as Reader					
Non-English	2.63	2.55	0.08	0.49	0.11
English	2.52	2.53	-0.01	0.94	-0.01
Student Identity					
Non-English	2.88	2.67	0.21**	0.02	0.34
English	2.83	2.74	0.09	0.19	0.13
Motivation in class					
Non-English	2.91	2.80	0.11	0.23	0.17
English	2.87	2.79	0.08	0.26	0.12
Reading Science					
Non-English	2.85	2.60	0.25**[A]	0.03	0.35
English	2.75	2.74	0.01	0.91	0.01

** statistically significant at the .05 level
* statistically significant at the .10 level

Integrated Learning Assessment: Table 10 shows means on the outcomes assessed via the Genetics ILA. Scores on the *Reading Process* rubric revealed significant differences between groups, with greater use among students in the intervention group of comprehension monitoring and problem-solving strategies that could build their understanding of the passage content. No other treatment/control differences on the ILA outcomes were detected.

Table 10. Student Integrated Learning Assessments by Treatment/Control Group

	Treatment	Control	Difference	p-value	Diff/SD
ILA Outcomes					
Writing Content	1.70	1.68	0.02	0.96	0.03
Writing Language	1.77	1.73	0.04	0.85	0.06
Reading Process	2.22	1.84	0.38**	0.03	0.42
Reading Text Summary	1.98	1.93	0.05	0.76	0.07
Metacognitive Content Understanding	3.05	3.07	-0.01	0.93	-0.01
Metacognitive Reading Comprehension	3.03	2.86	0.17	0.34	0.11

State Standardized Test Scores: To examine potential program impacts on student performance in biology and reading comprehension, we examine treatment/control differences state mandated criterion-referenced test scores. As described above, two types of test score data were collected - linked, longitudinal test score data for students for we had obtained parental consent and anonymous, unlinked, cross-sectional data student for students for whom we did not obtain parental consent. To account for treatment/control group non-equivalence in the sample retained, all analyses include controls for student and teacher characteristics measured prior to the intervention. **Table 11** shows the results based on both sets of test score data. For the longitudinal test data, students in treatment schools exhibited similar 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 .19 (Hill, Bloom, Black, & Lipsey, 2008). This indicates that students in the intervention classes were about one year ahead of their counterparts in control classes at the end of the study. Thus, there is 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. However, this inference should be tempered by that 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. It is only after we control for group differences in these pre-intervention characteristics do the program impacts on test scores become statistically significant.

Table 11. State Standardized Test Scores by Treatment/Control Groups.

	Treatment	Control	Difference	p-value	Diff/SD
2006/07 Test Scores (Longitudinal)					
ELA CST (std)	-0.02	-0.05	0.03	0.74	0.03
Reading Comprehension CST (std)	0.05	0.01	0.04	0.60	0.04
Biology CST scale score	338.07	330.10	7.97	0.27	0.17
2006/07 Test Scores (Cross-sectional)					
ELA CST (std)	0.10	-0.13	0.23**	0.04	0.23
Reading Comprehension CST (std)	0.10	-0.14	0.24**	0.02	0.24
Biology CST	336.52	322.91	13.61**	0.01	0.28

Notes. Data are based on regression-adjusted multilevel regression models that include randomization strata, racial/ethnic status indicator variables (Latino, Asian, African American, Other), EL 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 2004/05 teacher-aggregated Biology and reading comprehension test scores. Cross-sectional models include controls for 2004/05 teacher-aggregated Biology and reading comprehension test scores. ELA and reading comprehension CST scores are standardized 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 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 ($p < .05$)

Demographic Subgroups

We also examined program impacts by student racial/ethnic status, English learner status, and gender. **Tables 12a** and **12b** show impacts by subgroup for the longitudinal and cross-sectional samples, respectively. For the longitudinal test data, students in treatment schools, regardless of subgroup, exhibited similar levels of performance on the state standardized assessments as their counterparts in control schools. The two exceptions are the statistically significant positive impacts on reading comprehension for white students ($p < .10$) and biology for students classified as "other" on race/ethnicity. With so many statistical tests, however, it is possible that these two results are due to chance factors alone.

For the cross-sectional test data (**Table 12b**), 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 ($p < 0.10$). Overall, the pattern of results based on the cross-sectional data suggest that the impacts are most consistent and robust for whites and English learner students.

Table 12a. State Standardized Test Scores by Subgroup and Treatment/Control Groups – Longitudinal Results.

	Treatment	Control	Difference	p-value	Diff/SD
ELA CST					
Race/Ethnicity					
African American	-0.179	-0.272	0.092	0.616	0.110
Asian	0.090	0.163	-0.073	0.659	-0.076
Latino	-0.094	-0.011	-0.083	0.559	-0.095
Other	0.036	-0.040	0.076	0.721	0.079
White	0.096	-0.100	0.196	0.186	0.196
ELL Status					
Non-ELL	0.009	-0.044	0.053	0.857	0.053
ELL	-0.063	-0.040	-0.023	0.621	-0.025
Gender					
Female	0.083	-0.008	0.091	0.877	0.092
Male	-0.106	-0.089	-0.017	0.415	-0.017
Reading Comprehension					
Race/Ethnicity					
African American	-0.059	-0.161	0.102	0.576	0.112
Asian	0.190	0.166	0.024	0.874	0.026
Latino	0.014	0.055	-0.041	0.723	-0.044
Other	-0.072	-0.022	-0.050	0.789	-0.046
White	0.178	-0.051	0.229*	0.080	0.230
ELL Status					
Non-ELL	0.048	-0.007	0.055	0.958	0.054
ELL	0.050	0.056	-0.006	0.501	-0.006
Gender					
Female	0.128	0.003	0.125	0.735	0.128
Male	-0.021	0.008	-0.030	0.180	-0.030
Biology					
Race/Ethnicity					
African American	332.509	322.524	9.985	0.304	0.254
Asian	338.643	344.987	-6.344	0.466	-0.127
Latino	332.616	331.217	1.399	0.850	0.035
Other	350.349	325.238	25.111**	0.025	0.601
White	344.400	332.038	12.362	0.113	0.237
ELL Status					
Non-ELL	338.241	330.098	8.143	0.259	0.162
ELL	338.239	330.010	8.229	0.202	0.195
Gender					
Female	337.566	328.018	9.548	0.405	0.201
Male	338.360	331.911	6.449	0.222	0.132

Notes. Data are based on regression-adjusted multilevel regression models. See note for Table 7 for further details.
*Treatment/control difference is statistically significant ($p < .10$)
**Treatment/control difference is statistically significant ($p < .05$)

Table 12b. State Standardized Test Scores by Subgroup and Treatment/Control Groups – Cross-sectional Results.

	Treatment	Control	Difference	p-value	Diff/SD
ELA CST					
Race/Ethnicity					
African American	-0.296	-0.370	0.074	0.603	0.079
Asian	0.204	0.167	0.037	0.785	0.039
Latino	-0.008	-0.184	0.177	0.143	0.188
Other	0.163	-0.297	0.460**	0.002	0.474
White	0.356	0.019	0.337**	0.008	0.328
ELL Status					
Non-ELL	0.211	-0.047	0.259	0.148	0.255
ELL	-0.079	-0.246	0.167**	0.021	0.178
Gender					
Female	0.202	0.017	0.185**	0.021	0.190
Male	-0.004	-0.263	0.259	0.103	0.256
Reading Comprehension					
Race/Ethnicity					
African American	-0.274	-0.430	0.156	0.246	0.163
Asian	0.222	0.152	0.070	0.586	0.078
Latino	0.010	-0.200	0.210*	0.074	0.213
Other	0.146	-0.176	0.322**	0.041	0.326
White	0.339	-0.055	0.393**	0.001	0.396
ELL Status					
Non-ELL	0.199	-0.067	0.266*	0.081	0.265
ELL	-0.048	-0.249	0.201**	0.014	0.206
Gender					
Female	0.209	-0.029	0.238**	0.025	0.246
Male	0.000	-0.247	0.247**	0.032	0.242
Biology					
Race/Ethnicity					
African American	323.511	313.194	10.317	0.139	0.242
Asian	344.353	335.873	8.480	0.214	0.178
Latino	331.741	319.456	12.285*	0.056	0.282
Other	329.033	316.929	12.104	0.101	0.260
White	349.437	328.524	20.913**	0.001	0.383
ELL Status					
Non-ELL	340.051	324.301	15.750	0.125	0.298
ELL	331.169	321.717	9.452**	0.006	0.227
Gender					
Female	336.848	322.903	13.945**	0.011	0.298
Male	336.126	322.867	13.259**	0.008	0.259

Notes. Data are based on regression-adjusted multilevel regression models. See note for **Table 7** for further details.

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

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

Relationship of Implementation to Student Academic Performance

To examine how the level of Research Apprenticeship-aligned implementation in classrooms and what types of implementation practices might be related to student learning, we examined the relationship between the level of implementation and student test scores. Because the design does not involve random assignment to different types of implementation regimes, these analyses are purely descriptive in nature, and should not be used to make causal inferences. Nonetheless, the results from these analyses may be useful for planning subsequent experimental research.

Table 13. Correlations Between Implementation Level and Test Scores

	English Language Arts		*Reading Comprehension*		*Biology*	
	r	p-val	r	p-val	r	p-val
Overall Implementation						
All	-0.038**	0.01	-0.068**	<0.01	-0.063**	<0.01
Treatment	0.081**	0.03	0.049*	0.04	-0.048	0.94
Control	-0.130**	<0.01	-0.156**	<0.01	-0.093**	<0.01
Science Reading Opportunities						
All	0.125**	<0.01	0.077**	<0.01	0.038**	<0.01
Treatment	0.190**	<0.01	0.147**	<0.01	0.086**	<0.01
Control	0.066	0.35	0.015	0.84	-0.003	0.75
Teacher Support for Student Reading						
All	-0.056	0.28	-0.080	0.24	-0.064**	<0.01
Treatment	0.050**	<0.01	0.030**	<0.01	-0.048*	0.03
Control	-0.144**	<0.01	-0.171**	<0.01	-0.108**	<0.01
Metacognitive Inquiry						
All	-0.006**	0.04	-0.034	0.22	-0.015	0.64
Treatment	0.133**	<0.01	0.088*	0.06	0.052	0.46
Control	-0.141	0.12	-0.173	0.44	-0.108	0.29
Comprehension Strategies						
All	-0.109**	<0.01	-0.133**	<0.01	-0.121**	<0.01
Treatment	0.037*	0.06	0.008	0.07	-0.069	0.55
Control	-0.247**	<0.01	-0.259**	<0.01	-0.201**	<0.01
Collaboration						
All	-0.013	0.26	-0.049	0.28	-0.017	0.64
Treatment	0.120	0.11	0.072**	0.03	-0.028	0.44
Control	-0.121	0.36	-0.151	0.22	-0.003	0.86

Notes: Cross-sectional test score data set.
*Correlation is statistically significant at the .05 level.
**Correlation is statistically significant at the .10 level.

Table 13 shows Pearson correlations of implementation level and student test scores, by type of test and sample. A perusal of the table indicates the following:

- The level of implementation tends to be *positively* related to test scores in treatment classrooms but *negatively* related to test scores in control classrooms. The fact that the level of implementation is only positively related with test scores in treatment classrooms suggests that some unmeasured quality of implementation induced by the intervention lead to gains in student performance.

- In the treatment group, the associations between implementation and student performance are strongest and most consistent for English language arts scores, and weakest and least consistent for biology.
- In the treatment group, providing science reading opportunities and metacognitive inquiry are more strongly and consistently associated with student test scores than the other areas of implementation assessed via teacher interviews.
- Although the level of implementation tends to be *positively* related to test scores in treatment classrooms, the correlations are fairly week.

Overall, the relationship between the level of implementation and student test scores are consistent with the findings

SUMMARY

The study reported here has made 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.

Student OTL surveys and ILA assessments provide initial evidence, or leading indicators, that these differences in teaching result 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 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. On Integrated Learning Assessments, intervention students demonstrated more active intellectual engagement with the text and use of specific problem-solving strategies, a changed approach to reading science texts compared to control students. In addition, state standardized assessments provide 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 is associated with student performance in expected ways in treatment classrooms. The results of the study thus present a positive picture with regards to the effectiveness of the Reading Apprenticeship framework for integrating academic literacy content with biology coursework and instructional practices.

However, several cautions should be raised in interpreting the results. The study utilized a design in which teachers were recruited and randomly assigned to treatment and control groups fully two 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 *two* consecutive academic years. The implication of this sequencing is that treatment teachers had one academic year to *practice* using the framework, while 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. Retaining teachers in the study for such an extended period was a challenge. Such designs, while acknowledging the importance of practice for teachers, expose studies 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 even in one case, death.

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, 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. These treatment/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 suggested above, particularly in those schools serving the most vulnerable populations of 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, challenge to support diverse student needs, and demand from policy-makers.

Even with these caveats, the experimental impact estimates reported here are based on models that control for group differences in pre-intervention 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 impact teaching and learning outcomes in high school biology. 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 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 in biology content learning. At a time when strategies for improving educational outcomes for underachieving students increasingly focus on making structural changes, increasing accountability, and redistributing effective teachers through incentives for working in the most challenging schools, this is highly significant, demonstrating that it is possible to improve outcomes for students by building existing teachers' capacity through well designed professional development interventions of this kind.

Finally, the study indicates directions for improving the Reading Apprenticeship professional development intervention itself and suggests some potential next steps for research. The outcomes suggest that some aspects of the framework are harder for teachers to implement than others. For example, few teachers were able to engage students in metacognitive inquiry – exploring the nature of thinking and learning with science and science texts explicitly as a part of ongoing learning. Teachers need more support in this area. And while intervention teachers were able to fundamentally change the nature and degree of support they provided around existing course texts, few were able to change the texts themselves to offer a greater variety, interest, range of challenge levels, or the like. The fact that reading opportunities were most highly correlated with biology test score outcomes makes this an important focus of further exploration and potential development.

In addition, 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 on-going 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. Clearly, 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/literacy instructional "dose" that students would receive. Studies in New Zealand, for instance, have demonstrated sustained acceleration to close the achievement gap among mainstream and minority populations is possible, with sustained effort at a school site (Mai & McNaughton, 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.

We believe this study has demonstrated the promise of taking a disciplinary approach to literacy instruction, showing through a rigorous, scientific study design 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 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 not only would integrating literacy and science instruction mitigate these unintended consequences of restricting students' access to vital content area learning, but would 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.

REFERENCES

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: Longman.

Aschbacher, P. R. (1999). Developing indicators of classroom practice to monitor and support school reform (CSE Tech. Rep. No. 513). Los Angeles: UCLA/CRESST.

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

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: Jossey-Bass, Inc.

Barton, P. (2003). Parsing the achievement gap: Baselines for tracking progress. Policy Information Report of the Educational Testing Service. Princeton, NJ: ETS. Downloaded from the web in January, 2004 at 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 Publishing.

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.

Borasi, R. & Seigel, M. (2000). *Reading counts*. NY: Teachers College Columbia University.

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: Erlbaum.

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

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.

California Department of Education. (2000). *Construction of California's 1999 School Characteristics Index and Similar Schools Ranks.* California Department of Education, Office of Policy and Evaluation.

Clare, L. (2000). Using teachers' assignments as an indicator of classroom practice (CSE Tech. Rep. No. 532). Los Angeles: UCLA/CRESST.

Cook, T.D. & Campbell, D.T. (1979) Quasi-Experimentation: Design and Analysis Issues in Field Settings. Chicago: Rand-McNally, 1979.

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

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.

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), pp. 202 – 224.Durkin, 1984

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

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* (Technical Report No. 1-C). Berkeley, CA: National Center for the Study of Writing.

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

Greenleaf, C., Schoenbach, R., Cziko, C., & Mueller, F. (2001). Apprenticing adolescents to academic literacy. *Harvard Educational Review*, *71*(1), 79-129. www.wested.org/stratlit/pubsPres/HER/p01green.htm

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

Greenleaf, C.L. & Katz, M. (2004). Ever newer ways to mean: Authoring pedagogical change in secondary subject-area classrooms. In S.W. Freedman & A. F. Ball (Eds.), *New literacies for new times: Bakhtinian perspectives on language literacy and learning for the 21st century*. Cambridge: 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.

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 Dorothy Strickland and Michael L. Kamil, (Eds.), Improving Reading Achievement through Professional Development, Temple University: Laboratory for Student Success.

Guskey, T. R. & Huberman, M. (1996). *Professional Development in Education: New Paradigms and Practices*. Columbia, NY: Teachers College Press.

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.

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership, 58(6),* 6-11. 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: 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):172-77.

Hillocks, G. Jr. (1995). *Teaching writing as reflective practice.* NY: Teachers College Press.

Hull, G. A., & Rose, M. (1989). Rethinking remediation: Toward a social-cognitive understanding of problematic reading and writing. *Written Communication, 8*, 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.

Knapp, M. S., & Turnbull, B. (1991). Better schools for the children in poverty: Alternatives to conventional wisdom. Berkeley, CA: McCutchan.

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

Lee, C. (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.

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

Lemke, J. L. "Teaching All the Languages of Science: Words, Symbols, Images, and Actions." http://www-personal.umich.edu/~jaylemke/papers/barcelon.htm. This is an invited conference address that may be published in whole or in part in *Metatemas*, a journal edited in Barcelona.

Little, J. (2001). Inside Teacher Community: Representations of Classroom Practice. Paper presented as an invited address at the bi-annual conference of the International Study Association on Teachers and Teaching, Faro, Portugal, September 21-25.

Matsumura, L. C. (2003). Teachers, assignments and student work: Opening a window on classroom practice (CSE Tech. Rep. No. 602). Los Angeles: UCLA/CRESST.

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

McMurrer, J. (2007) *Choices, Changes, and Challenges: Curriculum and Instruction in the NCLB Era* Washington, DC: CEP

Miller, R. G. (2004). Creating thoughtful and expressive teachers and students using Read-Write activities: Final Report to NSF: IERI.

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), pp. 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.

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: U.S. Government Printing Office.

Murray, D. M. (1998) *Design and Analysis of Group Randomized Trials*. New York: Oxford University Press.

National Assessment of Educational Progress (2006). *The Nation's Report Card: Science 2005*. Washington, DC: NCES Retrieved 7/18/08 http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2006466

National Assessment of Educational Progress (2007). *The Nation's Report Card: Reading 2007*. Washington, DC: NCES Retrieved 7/18/08 http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007496

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

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

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

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). NY: Teacher's College Press.

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

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

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

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

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

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: 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.

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

Snow, C., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, D.C.: National Academy Press.

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

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

Wineburg, S. S. (1991). Historical Problem Solving: A Study of Cognitive Processes Used in the Evaluation of Documentary and Pictorial Evidence. *Journal of Educational Psychology* 83, pp. 73-87.

Appendix A – Professional Development Schedule

READING APPRENTICESHIP IN SCIENCE, SUMMER 2005

Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**
Introduction to the study **Model Lesson** Personal Science Reading Histories, Capturing Reading Processes and Reading Strategies Lists **Reading Process Analysis** Think Aloud with Acids and Bases	**Assessing and Re-Teaching** **Model Lesson** Extensive Reading building schema and Talking to the Text. Evolution Text set **Reading Process Analysis** Text and Task Analysis	**Assessing and Re-Teaching** **Professional Reading** What do Reading Apprenticeship Classrooms look like? **Student Literacy Case** The Pedagogy of Equity. What supports do struggling students need?	**Assessing and Re-Teaching** **Reading Process Analysis** Cell Theory text set. Thinking Aloud with visual models **Model lesson** Reading visuals, in science text. **Teaching Toolbox** Chunking, accessing prior knowledge. Repeated Readings	**Professional Reading** **Assessing and Re-Teaching** Got and Need **Supporting Classroom Application** Teachers draft an implementation plan for the first six weeks of school

Model Lesson KWL and Think Aloud: acid /base text set **Classroom video case** Introducing Think Aloud in Intro to Chemistry. **Teaching Toolbox** Scaffolding the metacognitive conversation **Assessing and Re-Teaching** Got and Need	**Student Literacy Case:** "Patterns of Evolution" in Modern Biology **Teaching Toolbox** Talking to the Text, building on the reading strategies list, Planning responsive instruction, Previewing and predicting from text structure. **Assessing and Re-Teaching** Got and Need	**Classroom video Case** What does a Reading Apprenticeship class look like in science? **Teaching Toolbox:** Supports for talk and inquiry **Supporting Classroom Application:** Plan supports for talk and reading **Assessing and Re-Teaching** Got and Need	**Supporting Classroom Application** Collaborative planning and mentoring Prepare a draft scope and sequence plan to share in the morning. **Assessing and Re-Teaching** Got and Need	**Supporting Classroom Application** Small groups meet, share plans, offer receive feedback Leave taking and logistics: Bring back a CERA sample! See you in January! **Assessing and Re-Teaching** Got and Need

NSF READING APPRENTICESHIP IN SCIENCE, WINTER 2006

Thursday	Friday
Assessing and Re-teaching Got and Need ***REFLECTING ON PRACTICE*** Assessing practice in the Dimensions of Reading Apprenticeship **Classroom Video Case** Acids and Bases Epilogue ***PROFESSIONAL READING*** building vocabulary and background knowledge	**Assessing and Re-teaching** Got and Need ***REFLECTING ON PRACTICE*** Analyzing student work and setting instructional goals **Professional Reading** Supporting Extensive Reading ***TEACHING TOOLBOX*** Classroom Libraries
TEACHING TOOLBOX word learning strategies in science ***MODEL LESSON*** Test as genre ***TEACHING TOOLBOX*** building schema: power standards, testing blueprints, test takers strategy list **Assessing and Re-teaching** Got and Need	**Supporting Classroom Application** Planning classroom libraries and next steps in metacognitive conversation **Assessing and Re-teaching** Got and Need

NSF READING APPRENTICESHIP IN SCIENCE, SUMMER 2006

Monday	**Tuesday**	**Wednesday**
Reflecting on practice Share lessons in trio's, reflect on Dimensions of Reading Apprenticeship **Reflecting on practice** Analyzing student work with the CERA rubric **Reading Process Analysis** **T**ext and Task Analysis of the CERA texts	**Assessing and Re-teaching** Got and Need **Reading Process Analysis** What kinds of questions do readers of science ask to make sense of science reading? **Model Lesson** Teaching students to be active questioners	**Assessing and Re-teaching** Got and Need **Reading Process Analysis** How do experienced readers of science clarify scientific text/language? **Model Lesson** Teaching word learning and clarifying strategies
Reflecting on practice Analyzing pre- and post instructional samples of student work and using CERA data to identify instructional needs and design responsive instruction. **Assessing and Re-teaching** Got and Need	**Teaching toolbox** ReQuest Question Answer Relationships Connecting to prior knowledge Monitoring conceptual change **Assessing and Re-teaching** Got and Need	**Supporting Classroom Application** Scope and Sequence: power standards as an equity tool Classroom Libraries Collaborative planning and conferring **Assessing and Re-teaching** Got and Need

Appendix B: Teacher Survey

Internal Consistency for Y1 Post-Institute Survey (across Control & Treatment Groups)

Construct	N	Item No.	Alpha
1. Science Reading Opportunities: Texts	77	q1a, q1b, q1c, q1d, q1e, q1f, q1g, q1h, q1i, q1j, q1k, q1l, q1m	.74
2. Science Reading Opportunities: Learning Structure	77	q2a, q2b, q2c, q2d, q2e, q2f	.69
3. Content	77	q2gr, q3b, q3dr, q6d, q6e	.53
4. Collaborative Activities: Teacher modeling, guidance, and support	77	q13b, q13d, q13h, q13i, q13g, q15d	.74
5. Collaborative Activities: Student Practice	77	q12a, q12d, q12e, q12f, q13c, q12g	.78
6. Megacognitive Inquiry: Teacher modeling, guidance, and support	77	q3c, q5a, q5b, q5c, q6b	.84
7. Megacognitive Inquiry: Student Practice	77	q4a, q4b, q4c, q4d, q4e, q6c, q6n	.90
8. Specific Comprehension Strategies: Teacher modeling, guidance, and support	77	q8a, q8d, q8e	.75
9. Specific Comprehension Strategies: Student Practice	77	q6f, q6l, q6g, q6h, q6i, q6j, q6k, q7b, q7d, q7e, q7f, q7g, q7j, q7i, q7q, q6a, q7l, q7m, q7n, q7o, q7k, q7r	.91
10. Negotiating Success: Instruction	77	q3a, q8b, q14c, q14e, q14f, q14g, q14h, q14i, q15f	.76
11. Negotiating Success: Assessment	77	q16b, q16c, q16d, q16h, q16er, q16i, q16j	.51
13. Teaching Philosophy: Reading	77	q18ar, q18br, q18c, q18dr, q18er, q18fr, q18gr, q18h, q18ir, q18jr, q18kr, q18lr, q18m, q18nr	.67
14. Teaching Philosophy: Learning	77	q19ar, q19br, q19c, q19dr, q19e, q19f, q19h, q19mr, q19n, q19or, q20e, q20k, q20lr	.67
15. Teaching Philosophy: Diversity	77	q20a, q20br, q20c, q20dr, q20f, q20ir, q20m, q20nr, q19ir, q19kr, q19lr	.47

Internal Consistency for Y2 Post-Institute Survey (across Control & Treatment Groups)

Construct	N	Item No.	Alpha
1. Science Reading Opportunities: Texts	58	q1a, q1b, q1c, q1d, q1e, q1f, q1g, q1h, q1i, q1j, q1k, q1l, q1m	.74 (.74)
2. Science Reading Opportunities: Learning Structure	59	q2a, q2b, q2c, q2d, q2e, q2f	.73 (.69)
3. Content	56	q2gr, q3b, q3dr, q6d, q6e	.10 (.53)
4. Collaborative Activities: Teacher modeling, guidance, and support	59	q13b, q13d, q13h, q13i, q13g, q15d	.80 (.74)
5. Collaborative Activities: Student Practice	57	q12a, q12d, q12e, q12f, q13c, q12g	.79 (.78)
6. Metacognitive Inquiry: Teacher modeling, guidance, and support	59	q3c, q5a, q5b, q5c, q6b	.74 (.84)
7. Metacognitive Inquiry: Student Practice	58	q4a, q4b, q4c, q4d, q4e, q6c, q6n	.87 (.90)
8. Specific Comprehension Strategies: Teacher modeling, guidance, and support	59	q8a, q8d, q8e	.77 (.75)
9. Specific Comprehension Strategies: Student Practice	53	q6f, q6l, q6g, q6h, q6i, q6j, q6k, q7b, q7d, q7e, q7f, q7g, q7j, q7i, q7q, q6a, q7l, q7m, q7n, q7o, q7k, q7r	.91 (.91)
10. Negotiating Success: Instruction	59	q3a, q8b, q14c, q14e, q14f, q14g, q14h, q14i, q15f	.69 (.76)
11. Negotiating Success: Assessment	57	q16b, q16c, q16d, q16h, q16er, q16i, q16j	.50 (.51)
12. Teaching Philosophy: Reading	58	q18ar, q18br, q18c, q18dr, q18er, q18fr, q18gr, q18h, q18ir, q18jr, q18kr, q18lr, q18m, q18nr	.31 (.67)
13. Teaching Philosophy: Learning	58	q19ar, q19br, q19c, q19dr, q19e, q19f, q19h, q19mr, q19n, q19or, q20e, q20k, q20lr	.61 (.67)
14. Teaching Philosophy: Diversity	56	q20a, q20br, q20c, q20dr, q20f, q20ir, q20m, q20nr, q19ir, q19kr, q19lr	.46 (.47)

Appendix C: Teacher Assignment Scoring Process and Inter-Rater Agreement

For each rubric, we used a 4-point scale (1=poor, 4=excellent) to rate the six dimensions of quality for each assignment. A total of twelve dimensions (six for content, six for literacy) were rated to evaluate the quality of teacher assignments. Each of the assignments was scored by at least two raters on 12 dimensions. Through discussion and using initial independent scores as a focus for these discussions, the raters established final consensus scores for all dimensions. Rubrics with more than a one point difference on at least one dimension were scored by a third rater. The Challenge dimension demonstrated the highest rater agreement in both rubrics with 57% of Content ratings and 61% of Literacy ratings receiving exact agreement (as illustrated in Table C.1). The Alignment of Goals and Task dimension demonstrated the lowest rater agreement in both rubrics with 44% of Content ratings and 41% of Literacy ratings with exact agreement. The final assignment ratings represent the consensus score across the raters.

Table C.1

Frequencies of Dimensions Receiving an Exact Agreement

Dimensions	**NSF Study N = 102**			
	Content		Literacy	
	n	%	*n*	%
Goals	57	56	47	46
Challenge	58	57	62	61
Support	55	54	48	47
Quality of Evaluation Criteria	51	50	56	55
Alignment Goals and Task	45	44	42	41
Alignment Goals and Evaluation Criteria	49	48	55	54

The frequencies of dimensions requiring third raters ranged from 2 – 12% (as illustrated in Table C.2).

Table 2

Frequencies of Dimensions Requiring a Third Rater

Dimensions	NSF Study N = 102			
	Content		Literacy	
	n	%	*n*	%
Goals	3	3	7	7
Challenge	2	2	4	4
Support	6	6	6	6
Quality of Evaluation Criteria	3	3	8	8
Alignment Goals and Task	6	6	12	12
Alignment Goals and Evaluation Criteria	4	4	4	4

Appendix D. Teacher Interview Report

The Interviews

Interviews were scheduled with each participating teacher in the study during the late spring of 2007. The interviews took place by phone and were recorded for analysis. Trained interviewers used an interview protocol and notetaker as they spoke by phone with the participating teachers. The interview protocol is attached. In total, 27 control teachers and 33 intervention teachers were interviewed using this process. Of the intervention teachers, 2 were categorized as intent to treat.

The Scores

Interviews were listened to repeatedly and trained scorers assigned scaled scores on five dimensions reflecting study constructs. Each construct and sub-construct was assigned a score ranging from 1-4 in accordance with the rubric (attached). Half point scores were utilized at the individual scorers discretion. Scoring sub-constructs allowed us to explore where particular teaching approaches showed up as differences between the intervention and control groups, reflecting the professional development and instructional framework of the intervention. Construct scores were assigned for each teacher by averaging sub-construct scores.

A sixth construct, Inquiry Science, was developed post hoc to the actual interviews. Therefore evidence of inquiry science classrooms was not necessarily explicit in the interviews. Because of this, during the initial scoring training cycle, scorers decided to score this sixth construct as a dichotomous outcome. Subjects were thus 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.

CONSTRUCT 1: SCIENCE READING OPPORTUNITIES

- Role of Reading
- Frequency of Reading
- Volume of Reading
- Breadth of Reading
- Accountability for Reading

CONSTRUCT 2: TEACHER SUPPORT FOR STUDENT EFFORTS TO COMPREHEND SCIENCE CONTENT FROM TEXT

- Where and How Reading and Comprehension Happens
- Nature of Teacher Support
- Student Practice
- Accountability/ Assessment of Content from Reading

CONSTRUCT 3: METACOGNITIVE INQUIRY INTO READING AND THINKING PROCESSES

Metacognitive Conversation
Teacher Instruction and Modeling of Metacognitive Processes, Routines, Tools and Strategies
Student Practice
Approach to Challenges
Accountability and Assessment of Metacognition

CONSTRUCT 4: SPECIFIC READING COMPREHENSION ROUTINES, TOOLS, STRATEGIES AND PROCESSES

Number of Strategies
Explicit Instruction and Modeling
Student Practice
Accountability and Assessment of Comprehension Strategies

CONSTRUCT 5: COLLABORATION

Frequency of Collaboration
Participation Structures and Routines that Support Collaboration
Focus of Collaboration
Teacher Support for Collaboration
Support for Equitable Participation
Accountability and Assessment

CONSTRUCT 6: SCIENCE INQUIRY (assigned 0/1; assign only if sufficient evidence)

Opportunities for Science Inquiry

The Scoring Process

Training

The four scorers began the process of scoring 57* teacher interviews by randomly selecting 3 interviews to score together. Each scorer listened to the recordings of each of these interviews and independently scored them on all constructs and sub-constructs. After all scorers had listened and scored the first interview, they met and discussed each of their individual scores for each sub-construct. Through this process of discussion of different scores and interpretations of the rubric, the rubric was revised to clarify distinctions. This process of individually listening and scoring the interviews, and then discussing those scores and the rubric was repeated 3 times, until scorers felt comfortable with the scoring process and the rubric.

**Although 60 interviews were conducted, 3 interviews proved impossible to score due to equipment failure and/or recording corruption.*

Reliability

In order to assure inter-scorer reliability, 10 interviews were randomly selected (5 treatment, 5 control) to be scored jointly by all scorers. The scorers then met approximately every 2 weeks throughout the scoring process to establish consensus on the scores of these interview subjects. Each meeting addressed the scoring of two interviews. This process served to enhance scorer confidence and insight into understanding and interpretation of the rubric. The scorers' goal was to come to within a one-point difference range on each of the overall construct scores. After each of these meetings, all four sets of scores were averaged across constructs and sub-constructs, with these averages reported as the final scores for these 10 interviews.

Interview Assignment and Scoring

Two of the scorers had conducted classroom observations in the classrooms of some of the interview subjects. In order to prevent these scorers from potentially biasing their scores based on what they had observed, rather than what they heard in the interview, the interviews of these subjects were assigned to the two scorers who had not observed in any classroom. Once those interviews were assigned, the remaining interviews were assigned to scorers at random. In addition to the 10 interviews that were scored jointly, the scorers were given the option to request a second score from the senior researcher on the team, should they feel uncertain about their scoring of any given interview. This option was utilized only once, by a scorer who discovered that an interview subject was known to her from another, independent, research project. In this case, both sets of scores were within one point of each other on all sub-constructs and the two sets of scores were averaged to report a final score for that interview.

Interview Analysis

Analysis of the interview data was conducted as t-tests comparing scores on constructs and sub-constructs for treatment and control teachers. Effect size (Cohen’s D) was calculated for each construct. Results (Table 1) show no significant differences between conditions for construct 1, Reading Opportunities, or any of the sub-constructs within that construct. Significant differences ($p<.001$) were found between groups for all four of the other constructs; Teacher Support for Student Efforts to Comprehend Science from Text (2), Metacognitive Inquiry Into Reading and Thinking Processes (3), Specific Reading Comprehension Routines, Tools, Strategies and Processes (4), and Collaboration (5). In addition, large effect sizes were also calculated for all 4 of these constructs. The effect size for construct 2 was the only one slightly below 1.0, while the effect size for constructs 3-5 were all greater than 1.0. The largest effect size (1.28) was found for construct 3, the construct designed to capture metacognitive inquiry teaching practices.

T-test Results for Means Comparisons of Biology Teacher Interview Construct and Sub-Construct Scores

	Treatment * (n=32)	**Control * (n=25)**	**Effect Size (Cohen's D)**	**p-value**
Construct 1: Reading Opportunities	2.84 (0.64)	2.64 (0.59)	0.33	0.22
Role of reading	2.96 (0.78)	2.80 (0.58)		0.37
Frequency of reading	3.13 (0.78)	2.92 (0.89)		0.36
Volume of reading	2.38 (0.84)	2.63 (0.60)		0.27
Breadth of reading	2.64 (0.92)	2.22 (0.79)		0.07
Accountability for reading	3.03 (1.05)	2.60 (0.78)		0.09
Construct 2: Teacher Support for Student Efforts to Comprehend Science from Text	2.92 (0.86)	2.10 (0.63)	**0.95**	**p<0.001**
Where and how reading and comprehension happens	2.94 (0.97)	2.02 (0.87)		p<0.001
Nature of teacher support	2.97 (0.92)	2.00 (0.14)		p<0.001
Student practice	3.19 (0.86)	2.52 (0.77)		p<0.01
Accountability/Assessment of content from reading	2.50 (1.13)	1.68 (0.71)		p<0.01
Construct 3: Metacognitive Inquiry Into Reading and Thinking Processes	2.55 (0.66)	1.56 (0.51)	**1.28**	**p<0.001**
Metacognitive conversation	2.47 (0.84)	1.52 (0.53)		p<0.001
Teacher instruction and modeling of metacognitive processes, routines, tools, and strategies	2.63 (0.78)	1.64 (0.64)		p<0.001
Student practice	2.72 (0.89)	1.46 (0.58)		p<0.001
Approach to challenges	2.77 (0.73)	1.86 (0.60)		p<0.001
Accountability and	2.34	1.44		p<0.001

assessment of metacognition	(1.00)	(0.73)		
Construct 4: Specific Reading Comprehension Routines, Tools, Strategies and Processes	2.98 (0.75)	2.02 (0.80)	**1.06**	**p<0.001**
Number of strategies	2.95 (0.86)	2.04 (0.79)		p<0.001
Explicit instruction and modeling	2.72 (0.78)	1.74 (0.77)		p<0.001
Student Practice	3.17 (0.89)	2.18 (0.95)		p<0.001
Accountability and assessment of comprehension routines	2.63 (0.95)	1.88 (0.77)		p<0.01
Construct 5: Collaboration	2.97 (0.62)	2.14 (0.44)	**1.21**	**p<0.001**
Frequency of collaboration	3.58 (0.62)	2.92 (0.81)		p=0.001
Participation Structures and routines that support collaboration	3.14 (0.71)	2.22 (0.61)		p<0.001
Focus of collaboration	3.27 (0.75)	2.28 (0.66)		p<0.001
Teacher support for collaboration	2.56 (0.86)	1.64 (0.45)		p<0.001
Support for equitable participation	2.66 (0.98)	1.70 (0.63)		p<0.001
Accountability and Assessment	2.63 (0.78)	1.72 (0.46)		p<0.001

**Standard Deviations appear in parentheses below means*

Teacher Interview Protocol, Spring 2007

Name of Teacher ____

Name of Interviewer ____

Date _________________________

Time ________________________

Introduction to Interview

This study focuses on literacy in the biology classroom and its potential impact on students' learning and achievement. With an accurate picture of what biology teachers are doing, we hope to be able to figure out if and how science reading instruction makes any difference for students. I will be asking you a series of questions about your classroom teaching, but please don't feel there are any wrong answers to these questions. I will also be asking you to focus on the one class from which we are collecting data for this study. My goal is to get as accurate a picture as possible of this class and your teaching.

[Turn tape recorder on.]

I will be audiotaping this interview so I can make sure I've captured your responses accurately. Is that okay with you?

Course Goals

1. I'd like to start by asking you to describe the focus of your biology instruction in this focus class. What are your goals for student learning in this class?

[ ] development of procedures/facts/information/vocabulary
[ ] covering standards (which ones?)
[ ] preparing students for state assessments
[ ] building interest in science
[ ] development of science concepts
[ ] development of literacy proficiencies
[ ] development of science thinking and reasoning
[ ] development of science investigation skills
[ ] building wise consumers of science/decision making citizens
[ ] developing dispositions for learning
[ ] building basic skills in science (behaviors, measurement, etc.)
[ ] other ________________________________

2. As you know, this study explores the impact of literacy instruction on student achievement and learning in biology. We define literacy quite broadly as the full range of reading, writing,

and discussion activities students use in biology. Are you working on developing student literacy in science in any way in the focus class?

Yes No [If not, skip to question #3]

What are your overall goals in this class for student's literacy learning?

[ ] access to ideas of science
[ ] strengthening particular reading/writing skills
[ ] building independence as learners of science
[ ] building science thinking
[ ] building science reading proficiency
[ ] building science writing proficiency
[ ] improving student expression of science ideas/discourse
[ ] willingness to pick up a text and try to make sense of it
[ ] enjoying reading science
[ ] building confidence in their ability to attack science texts and gain understanding
[ ] helping students work with challenging texts
[ ] holding students responsible for reading lab instructions
[ ] other: ____

Instructional Sequence, Instructional Approaches, and Daily Routines

3. Can you give me a sense of what you've done from the beginning of the year to now in this class to work on these goals or build these skills?

Has this changed in any way from what you may have planned at the start of the school year **[for example, in the professional development institute]**? If so, how, and why did you change your plans?

4. You've said **[reference anything the teacher said in #2 or #3 with respect to efforts to assist students with literacy learning goals]**. What if any instructional approaches or strategies do you use to help students in this class with the science reading?

Probe any responses for clarity: Can you give me an example of that?

5. [**If the teacher has not described modeling as a strategy, ask**] How often, if at all, do you model or demonstrate how to go about reading or writing tasks in this class?

never ⇔ rarely ⇔ sometimes ⇔ often ⇔ all or almost all lessons

Would you explain that a bit?

[**If teacher does model**] How do you go about modeling or demonstrating literacy tasks? Could you give me an example?

To what extent would you say your modeling typically focuses on explaining your thinking and reasoning processes to show students how to think more scientifically, versus focusing primarily on procedures to make sure students do the task correctly?

procedures ⇔ reasoning processes

6. You've given me a sense of your big picture goals and teaching approaches. Now could you describe your daily classroom routines? How does a typical day go in this class?

[**If the teacher has not already indicated, ask:**] Is reading or writing a typical part of the day? If so, in what way? How do [**reference teacher's literacy goals from #2 and approaches from #3 or 4**] fit into a typical day?

Teacher Beliefs and Practices about Reading in Science

7. Now I'd like to get a more detailed picture of the reading students do for this class. Do you assign science reading to this class either in class or for homework?

[ ] yes [ ] no [ ] sometimes

If so, why? What role do you see the text playing in your students' science learning?

[ ] Building interest in the topic
[ ] Providing background (introducing a topic, concepts, or vocabulary)
[ ] Delivering content (providing information)
[ ] Making connections to real world science
[ ] Modeling science inquiry
[ ] Modeling literacy practices
[ ] Providing data for students to interpret (secondhand inquiry)
[ ] Supporting students' firsthand science inquiry
[ ] Explicating or developing concepts
[ ] Providing multiple representations of ideas and concepts
[ ] Providing contexts for inquiry and problem solving
[ ] Supporting students' research
[ ] Building students' academic reading proficiency
[ ] Building students' science reasoning processes
[ ] Practicing reading
[ ] Seeing reading and reading materials as a resource for learning
[ ] Other: ________________________________

8. How important do you believe it is for students to understand the science reading materials of your biology curriculum?

not important ⇔ somewhat important ⇔ important ⇔ very important

Why do you say this?

9. How do you use the textbook in this class, if at all?

What is the textbook title and publisher (get publication date, edition)?

____

10. You've talked about **[reference any reading materials the teacher has mentioned].** Do you assign any reading materials other than the textbook? If so, what materials? **[Probe for starred items.]**

[ ] laboratory procedures *
[ ] science news articles *
[ ] journal articles *
[ ] teacher made lecture notes, chalk boards, white boards *
[ ] internet *
[ ] computer software *
[ ] student generated work (specify, for ex: white boards, journals, lab reports) *
[ ] data sets *
[ ] trade books
[ ] science biographies
[ ] nonfiction science books on specific topics
[ ] science related fiction
[ ] powerpoint slides accompanying lectures
[ ] other (specify):

Do the students in this class read graphs, tables, models or science illustrations? If so, how do they typically work with these materials?

[**If not already answered above**] Do students in this class read multiple texts on a specific topic? If so, what might these multiple texts include? For what purpose?

[**If not already answered above**] Do students in this class do any extended reading (chapters, articles, books, other lengthy texts)? If so, please give an example.

What about any self-chosen reading? If so, what? When? For what purpose?

How, if at all, have you used the classroom library that was offered to you as a participant in this study?

11. <u>How often</u> do you assign the students in this class to read science materials, whether for homework or during class? **[You can help teachers quantify frequencies by translating them as follows:** Never, Rarely (e.g. a few times a year), Sometimes (e.g., a few times a month), Often (e.g., once or twice a week), All or almost all lessons]

never ⇔ rarely ⇔ sometimes ⇔ often ⇔ all or almost all lessons

<u>Where</u> do students typically complete the reading (**in class, out of class**)?

<u>How</u> do they typically complete the reading (**independently, collaboratively**)?

What do you typically have students do before, during and after the reading?

You've said that students typically complete readings for this class (**for homework, during class**). How frequently do students in this class read during class time?

never ⇔ rarely ⇔ sometimes ⇔ often ⇔ all or almost all lessons

Why **do you/don't you** choose to have students work on science reading in class?

Has the frequency of students' reading during class changed at all over the course of the year? If so, how and why has this changed?

12. How, if at all, do you assess or grade student reading of science materials in this class?

What are you looking for when you assess their reading? What is the focus of your assessment of students' science reading in this class?
[ ] completion/compliance
[ ] participation/effort
[ ] comprehension
[ ] evidence of student biology learning
[ ] evidence of reading strategies and/or processes
[ ] depth of discussion
[ ] degree of intellectual rigor
[ ] other ______________________________________

Has the focus of your reading assessment changed at all during the year? If so, how and why?

Description of Instruction

13. You've talked about **[reference anything the teacher said about collaboration or classroom interactions]**. I'm going to ask some questions to get a more detailed picture of the kinds of interactions that are typical in this class.

How often do students work together on class assignments? What about on reading assignments?

never ⇔ rarely ⇔ sometimes ⇔ often ⇔ all or almost all lessons

Can you describe the nature of this collaboration—for example, do students work together informally with friends, or with assigned partners or groups? What happens in the groups?

What about when they are working on reading?

14. How has this collaboration gone in this class this year? How satisfied are you with students' discussions and group work?

[ ] not at all satisfied **[ ] somewhat satisfied** **[ ] very satisfied**

What kinds of things have you seen to make you say that?

Have you done anything in particular to encourage and support collaboration in this class?

[ ] Yes **[ ] No**

If so, what have you done?

15. What is your role during group work? How do you interact with individuals or groups?

[ ] monitor small group or individual work for on task behavior
[ ] monitor small group or individual work to assess understanding
[ ] step in to facilitate group processes as needed
[ ] support students' inquiry, e.g., by asking open-ended questions, encouraging students to work to solve problems or pursue own questions
[ ] check on students' learning, understanding
[ ] respond to student requests for help by providing information or answers to their questions
[ ] other ____

Can you give me a picture of what that looks like?

Could you give me an example of a typical interaction you might have with a small group?

16. Now I'd like to ask a few questions about how you handle challenging ideas and materials in this class. To what extent do you typically have students spend time grappling with concepts or ideas in this class, or do you generally explain things you want students to get?

students grapple ⇔ teacher explains

What about with challenging reading?

students grapple with reading ⇔ teacher explains reading

Why?

17. 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?

never ⇔ rarely ⇔ sometimes ⇔ often ⇔ all or almost all lessons

Why do you have students spend time in this way? Can you tell me more about what this looks like?

18. In addition to discussing the biology content, how often, if at all, do you have students in this class discuss their thinking and problem solving processes? For example, sharing *how* they learn or *how* they approach class assignments?\

never ⇔ rarely ⇔ sometimes ⇔ often ⇔ all or almost all lessons

Can you give me a picture of how they do that?

19. Is there anything else you'd like to tell me about your biology class, your instruction, or your students?

Thank you so much for taking this time to describe your students and your biology instruction. I appreciate having a good picture of your classroom and know this interview will contribute to what we learn from the study.

NSF Teacher Interview Scoring Rubric
January 17, 2008

CONSTRUCT 1: SCIENCE READING OPPORTUNITIES				
	4	**3**	**2**	**1**
Role of Reading	Reading plays a central role in science learning; teacher emphasizes role of reading—e.g., to provide access to ideas of science, build science thinking, model science inquiry and literacy practices. Teacher stance characterized by "always literacy and as much inquiry as possible" Teacher claims reading is critical for science learning Teacher has explicit, elaborated science reading goals Reciprocity between reading and hands-on science—e.g., students work back and forth between texts and lab investigation. Little apparent boundary between reading and science investigation	Reading plays a supporting role in science learning; teacher emphasizes complementary role of reading—e.g., to build interest in topic, make connections to real world, support students' firsthand science inquiry Teacher claims reading is important for science learning Teacher has explicit, science reading goals. Goals may be somewhat general or limited Explicit connections are made between reading and other science activities/assignments—e.g., students read textbook to learn about lab	Reading plays a supplementary role in science learning; teacher assigns reading for non-essential learning— e.g., extra credit, bonus points Teacher claims reading is somewhat important for science learning Teacher may not have explicit science reading goals or may have vague, literacy goals (e.g., "to read better") Some articulation between reading and hands-on science—e.g., reading assignments may generally correlate with unit topic—but little explicit connection is made between reading and classroom work	Reading plays little role in science learning; regardless of whether or not teacher assigns reading, reading carries little weight in science learning Teacher diminishes role of reading or claims reading is unimportant for his/her students compared with other science learning activities Teacher has no science reading goals No noticeable articulation between reading and other activities/assignments, no explicit connection between reading and classroom work
Frequency of Reading	Reading is assigned with the understanding that students are to read it in every or nearly every lesson	Reading is assigned with the understanding that students are to read it in at least half but less than every lesson (i.e., two or three times a week)	Reading is assigned with the understanding that students are to read it in less than half of all lessons (i.e., once or twice a week)	Reading is rarely or never assigned with the understanding that students are to read it (i.e., once or less than once a week)
Volume of Reading	Teacher assigns large volumes of text: equivalent of textbook (500 pages/year) plus 10 additional pages a week from articles and other supplementary texts	Teacher assigns equivalent of textbook (500 pages/year)	Teacher assigns equivalent of half the textbook (250 pages/year). Teacher may typically assign excerpts from textbook, other texts	Teacher assigns equivalent of one quarter of the textbook or less (125 pages/year). Materials may be limited to teacher notes, handouts
Breadth of Reading	Teacher continuously assigns a wide range of instructional genres/text types (five or more) serving a variety of purposes	Teacher regularly assigns a number of instructional genres/text types (up to four) serving a variety of purposes. Teacher may utilize the range of genres/text types found in traditional classroom materials (textbook and lab materials) and regularly supplement with an additional text type (e.g., newspaper, magazine and internet articles)	Teacher assigns three instructional genres/text types—e.g., textbook, lab materials and occasional supplementary text	Teacher assigns fewer than three instructional genres/text types—e.g., textbook and lab materials with rare or no supplementary text

Accountability for Reading	Students are often held accountable for reading assignments. Students cannot meet class expectations without reading	Students are sometimes held accountable for reading assignments. Students cannot meet class expectations without reading	Students are occasionally held accountable for reading assignments. Teacher often reinforces important content from reading verbally (e.g., through lectures). Students can meet most class expectations without reading	Students rarely or never held accountable for reading assignments. Teacher may not collect reading assignments, or may grade on completion only (e.g., a check or stamp). Important content is almost always presented verbally (e.g., through lectures). Students can meet class expectations without reading

CONSTRUCT 2: TEACHER SUPPORT FOR STUDENT EFFORTS TO COMPREHEND SCIENCE CONTENT FROM TEXT				
	4	**3**	**2**	**1**
Where and How Reading and Comprehension Happens	Reading and comprehending, or comprehending, of science content from text frequently happens in class with social support (every or nearly every lesson). May be a shift toward increased reading outside the classroom as the year progresses, but supported in-class comprehension of content is ongoing	Reading and/or comprehending, or comprehending, of science content from text sometimes happens in class with social support. Supported in-class reading occurs regularly but may be limited to specific times/purposes, e.g., “Metacognitive Mondays” or when introducing/practicing a new reading comprehension strategy or routine	Reading and/or comprehending, or comprehending, of science content from text occasionally happens in class with social support. Supported in-class reading is sporadic or short-lived	Reading and/or comprehending, or comprehending, of science content from text rarely or never happens in class with social support. In-class reading, if any, is independent and unsupported
Nature of Teacher Support	Teacher support frequently fosters student agency for understanding science content from text. Teacher promotes and orchestrates student problem-solving and meaning-making	Teacher support sometimes fosters student agency for understanding science content from text. Teacher provides some support for student problem-solving and meaning-making, but often hints, helps and solves comprehension problems	Teacher support occasionally fosters student agency for understanding science content from text. Teacher may answer student questions about science content or lead “discussions” in which teacher fishes for right answers (I-R-E pattern)	Teacher rarely or never supports student agency for understanding science content from text. Teacher may “read for” students by delivering science content verbally, or ignore readings altogether
Student Practice	Students frequently do the work of reading and comprehending science content from text	Students sometimes do the work of reading and comprehending science content from text	Students occasionally do the work of reading and comprehending science content from text	Students rarely or never do the work of reading and comprehending science content from text
Accountability/ Assessment of Content from Reading	Teacher often assesses students’ understanding as they read and make sense of science content and uses it to guide literacy instruction and support for comprehending science content from text	Teacher sometimes assesses students’ understanding as they read and make sense of science content and uses it to guide literacy instruction and support for comprehending science content from text	Teacher occasionally assesses students’ understanding as they read and make sense of science content and uses it to guide literacy instruction and support for comprehending science content from text. Use of formative assessment may be reactive, e.g., teacher provides short lesson on reading graphs after class fails test	Teacher rarely or never assesses students’ understanding as they read and make sense of science content and uses it to guide literacy instruction and support for comprehending science content from text, or grades based on completion only

CONSTRUCT 3: METACOGNITIVE INQUIRY INTO READING AND THINKING PROCESSES				
	4	**3**	**2**	**1**
Metacognitive Conversation	Reading frequently involves noticing, sharing and problem-solving confusions, reading and thinking processes and sense-making	Reading sometimes involves noticing, sharing and problem-solving confusions, reading and thinking processes and sense-making (1-2 times a week). May be more frequent, but limited, e.g., confusions and understanding may focus primarily on right answers, rather than reading and thinking processes	Reading occasionally involves noticing, sharing and problem-solving confusions, reading and thinking processes and sense-making—e.g., sporadically or at the beginning of the year	Little or no opportunity for students to notice, share or problem-solve confusions, reading and thinking processes and sense-making
	Metacognitive conversation about reading and thinking processes takes center-stage in an ongoing metacognitive conversation	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 occasionally engages students in metacognitive conversation about reading and thinking processes—e.g., at the beginning of the year only, of narrow focus, through a single routine or tool, used sporadically.	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 often teaches and models reading and thinking processes, routines and strategies that support students to become self-monitoring and self-governing readers of science	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 (once a week or so)	Teacher occasionally teaches and models reading and thinking processes, routines and strategies—e.g., sporadically or at the beginning of the year)	Teacher does not teach or model reading and thinking processes, routines or strategies
	Modeling and demonstration generally emphasizes reading and thinking processes rather than procedures	Modeling and demonstration sometimes emphasizes reading and thinking processes and sometimes emphasizes procedures and correctness	Modeling and demonstration may occasionally focus on reading and thinking processes, but generally emphasizes procedures and correctness	Teacher modeling and demonstration, if any, almost always focuses on procedures and correctness rather than reading and thinking processes
Student Practice	Students often practice metacognitive reading routines, tools and strategies (most lessons)	Students sometimes practice metacognitive reading routines, tools and strategies (once or twice a week), or student practice, though more frequent, focused on answers rather than reading and thinking processes—e.g., double entry reading logs focused on facts/examples	Students occasionally practice metacognitive reading routines, tools, strategies and conversations (a few times a month)	Students rarely or never practice metacognitive reading routines, tools, strategies and conversations (once a month or less)
Approach to Challenges	Teacher usually encourages and supports students to grapple with challenging texts, tasks and concepts	Teacher sometimes encourages and supports students to grapple with challenging texts, tasks and concepts, or grappling, though frequent, may involve limited teacher support in how to approach challenging materials	Teacher occasionally has students grapple with challenging texts, tasks and concepts, albeit with little teacher support	Teacher rarely or never has students grapple with challenging texts, tasks and concepts

Accountability and Assessment of Metacognition	Teacher frequently collects and/or assesses students' reading and thinking processes, using 2 or more measures (e.g., reading logs, annotations, student talk). Focus of assessment may shift across year, e.g., from focus on comprehension monitoring and strategies use to evidence of scientific reading and thinking, but teacher continues to assess reading and thinking process as well as content	Teacher sometimes collects and/or assesses students' reading and thinking processes, e.g., after teaching a new routine or strategy. Focus of assessment may shift across year from focus on reading processes to content/correctness	Teacher occasionally collects and/or assesses students' reading and thinking processes. Assessment may be sporadic or short-term	Teacher rarely or never collects and/ or assesses students' reading and thinking processes

CONSTRUCT 4: SPECIFIC READING COMPREHENSION ROUTINES, TOOLS, STRATEGIES AND PROCESSES				
	4	**3**	**2**	**1**
Number of Strategies	Teacher introduces an appropriate number of high leverage reading comprehension supporting strategies (e.g., building and activating schema, RT strategies of clarifying, questioning, connecting and predicting) which are used over and over in multiple contexts	Teacher introduces a somewhat restricted number of comprehension supporting strategies which are used repeatedly in different contexts	Teacher introduces one or two comprehension supporting strategies (e.g., Cornell notes) or presents a smorgasbord of strategies, too few to solve reading comprehension problems, or too many to develop expertise	Teacher introduces no comprehension supporting strategies
Explicit Instruction and Modeling	Teacher provides explicit instruction and models flexible use of comprehension supporting routines, tools and strategies on a regular and ongoing basis. Instruction addresses why and when of strategy use, as well as procedures	Teacher sometimes provides explicit instruction and models comprehension supporting strategies. Instruction may be limited to introduction of new routines, tools and strategies or may be sporadic. Instruction addresses why and when of strategy use, as well as procedures	Teacher occasionally provides explicit instruction and models comprehension supporting routines, tools and strategies. Teacher may frontload instruction at the beginning of the year without additional instruction. Instruction may focus on procedures rather than why or when	Teacher rarely or never provides explicit instruction or models comprehension supporting routines, tools and strategies (once or twice)
Student Practice	Teacher assigns comprehension supporting routines, tools and strategies as an integral part of science reading. Students often use comprehension strategies (e.g., questioning, summarizing) to support them in making meaning while they read, rather than summatively (i.e., to demonstrate comprehension)	Teacher sometimes assigns comprehension supporting routines, tools and strategies with reading. Students sometimes use comprehension strategies to support them in making meaning while they read, although they may sometimes use them summatively (i.e., to demonstrate comprehension)	Teacher occasionally assigns comprehension supporting routines, tools and strategies with reading. Strategies may not be integral to reading but taught and practiced for their own sake, or may be used after reading to demonstrate comprehension rather than during reading to support meaning making	Teacher rarely or never assigns comprehension supporting routines, tools and strategies with reading
Accountability and Assessment of Comprehension Strategies	Teacher frequently assesses and otherwise holds students accountable for use of comprehension supporting strategies on an ongoing basis, e.g., through RT, reading logs Teacher routinely monitors use of comprehension strategies and provides additional support and reteaching on an ongoing basis	Teacher sometimes assesses or otherwise holds students accountable for use of comprehension supporting strategies, e.g., on “Metacognitive Mondays” or for a period after introducing each new strategy or tool Teacher sometimes monitors use of comprehension strategies and provides additional support and reteaching	Teacher occasionally assesses or otherwise holds students accountable for use of comprehension supporting strategies, e.g., at the beginning of the year Teacher occasionally monitors use of comprehension strategies and provides additionally support and reteaching	Teacher rarely or never assesses or otherwise holds students accountable for use of comprehension strategies Teacher rarely or never monitors use of comprehension strategies

CONSTRUCT 5: COLLABORATION				
	4	**3**	**2**	**1**
Frequency of Collaboration	Students frequently work collaboratively	Students sometimes work collaboratively	Students occasionally work collaboratively	Students rarely or never work collaboratively. Teacher may oppose or discourage group work
Participation Structures and Routines that Support Collaboration	Teacher establishes a small number of well-established, smoothly running structures/routines used over and over to support collaborative meaning making (e.g., pair-share, RT, partner reading)	Teacher establishes some structures/routines that support collaborative meaning making. Structures may be well-established but somewhat restricted in number or purpose (e.g., pair-share only), or teacher may try multiple structures infrequently	Collaboration is largely incidental and informal, rather than structured (e.g., "I let them talk with a neighbor if they are not too loud"). Teacher may have students work in groups, but with little structure or routine	Collaboration, if and when it occurs, is driven by pragmatic reasons, resource availability (e.g., textbook availability, lab space, time constraints)
Focus of Collaboration	Collaboration/talk always or frequently focuses on academic learning goals, supports students' intellectual and academic development. Social aspects of collaboration leveraged to support students' intellectual and academic development	Collaboration/talk focuses partly on academic learning goals and intellectual and academic development and partly on social goals, e.g., "students like to work in groups"	Collaboration only occasionally focuses on intellectual and academic development. Teacher focuses primarily on social benefits and challenges of collaboration/talk, rather than on academic learning goals	Teacher focuses on difficulties and pitfalls of collaboration/talk
Teacher Support for Collaboration	Teacher frequently teaches, models and supports collaborative processes, e.g., through itinerant mentoring	Teacher sometimes teaches. models and supports collaborative processes	Teacher occasionally teaches, models and supports collaborative processes	Teacher rarely or never teaches, models and supports collaborative processes
Support for Equitable Participation	Teacher establishes a number of effective routines, structures and interventions to ensure participation of all students(e.g., popsicle sticks, itinerant mentoring, rotating roles, turn-taking routines)	Teacher implements one or more routines, structures and interventions to promote participation of all students, but is only partially successful	Teacher verbally encourages but does not establish routines, structures and interventions to further support and ensure participation by all students	Little or no attempt to support equitable participation in group. Group work inequitably distributed, e.g., more competent students may take over work, or teacher may accept that some "deadbeat" students will not participate
Accountability and Assessment	Teacher frequently monitors group work for evidence of collaborative processes	Teacher sometimes monitors/assesses group work for collaborative processes. Emphasis may be on procedures (e.g., steps in task, division of task) rather than collaborative meaning making	Teacher occasionally monitors/assesses collaborative processes as well as science content. Focus on compliance (on task behavior) rather than collaborative meaning-making	Teacher rarely or never monitors/assesses group work

	Teacher often holds students accountable for contributions to the group and holds the group responsible for the learning of individual members, e.g., through self-assessment/reflection, group grade, sharing or presentation/spokesperson role	Teacher sometimes holds students accountable for their contributions to the group and/or holds the group responsible for the learning of individual members. Weight may be on individual contribution, product	Teacher occasionally holds students accountable for their contributions to the group and/or holds the group responsible for the learning of individual members. Students generally assessed on individual work only	Teacher rarely or never holds students accountable for their contributions to the group or holds the group responsible for the learning of individual members. Students always or nearly always assessed on individual work only

CONSTRUCT 6: SCIENCE INQUIRY (assigned 0/1; assign only if sufficient evidence)				
	4	**3**	**2**	**1**
Opportunities for Science Inquiry	**Science instruction and classroom interactions frequently focus on science inquiry—i.e., gathering and interpreting evidence. The following characteristics are generally descriptive of the class:**	Science instruction and classroom interactions sometimes focus on science inquiry—i.e., gathering and interpreting evidence—and other times focuses on acquiring facts. The following characteristics are generally descriptive of the class:	Science instruction and classroom interactions occasionally focus on science inquiry—i.e., gathering and interpreting evidence—and generally focus on acquiring ready-made knowledge. The following characteristics are generally descriptive of the class:	Science instruction and classroom interactions rarely or never focus on science inquiry and almost always focus on acquiring ready-made knowledge. The following characteristics are generally descriptive of the class:
	Teacher instruction, modeling and demonstration emphasizes scientific thinking and processes	Teacher instruction, modeling and demonstration sometimes focuses on scientific thinking and processes and sometimes focuses on procedures and correctness	Teacher instruction, modeling and demonstration occasionally focuses on scientific thinking and processes; however, modeling and demonstration generally focus on procedures and correctness rather than thinking	Teacher rarely or never teaches, models or demonstrates scientific thinking and processes; modeling and demonstration, if any, focus on procedures and correctness rather than thinking
	Students frequently practice science inquiry (i.e., activities that focus on student thinking, investigation and problem-solving, including but not limited to “inquiry labs”)	Students sometimes practice science inquiry (i.e., activities that focus on student thinking, investigation and problem-solving, including but not limited to “inquiry labs”) and sometimes do activities that focus on acquiring “ready-made” knowledge (e.g., worksheets, “cookbook labs”) .	Students occasionally practice science inquiry (i.e., activities that focus on student thinking, investigation or problem-solving, including but not limited to “inquiry labs”). Activities generally focus on acquiring “ready-made” knowledge (e.g., worksheets, “cookbook labs”)	Students rarely or never practice science inquiry. Activities almost always involve acquiring “ready-made” knowledge (e.g., worksheets, “cookbook labs”)

Appendix E: Summary of Constructs and Reliability Studies of Opportunity to Learn Survey

Appendix Table E-1. Student *Opportunity to Learn Survey Constructs*

		Alpha
Item	(1) Class Emphasis on Reading in Biology	0.80
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…teacher's mind while the teacher reads science mater.	
11	Encouraged students to borrow one another's ideas.	
	(2) Frequency of Student Integration of Biology and Literacy Activity	0.77
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 (e.g., headings,..).	
	(3) Perceived Course Consequences on Identifying as a Reader	0.74
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	0.80
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	0.84
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.	
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	0.90
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.	

Appendix F: Integrated Learning Assessment Scoring Process and Inter-Rater Reliability

Sample Size

Teacher

Of the sixty-three teachers participating in this study, 28 submitted ILAs representing 27 public high schools in California. Eighteen of these teachers were women and 10 were men. Their length of teaching experience ranged from 3 to 36 years, with an average of 5.4 years.

Student

A total of 701 students were administered the ILA by 28 participant teachers. Of the 701 students, 309 (44%) were female, 296 (42%) were male, and 96 (14%) did not report gender. For the statistical analysis, students for whom we did not have teacher level data were excluded in the final analysis. As a result, a total of 676 students were included in the final analysis.

The Scoring Session

A total of 701 student responses were evaluated during the 7-day scoring session. To minimize rater bias, all identifying information (student names, teacher names, and school names) was removed from the student papers. Responses were randomly distributed and divided into packets containing 20 responses each.

All raters underwent intensive training to introduce and practice scoring procedures, address questions, and ensure that the scoring rubrics were clear. Training on days 1 and 2 focused on the Content and Language rubrics. All student writing responses were then scored by three different raters to achieve greater consistency. The final scores for student responses represent the arithmetic mean of the three raters' scores. After the writing section of the ILA was scored, training on days 5 and 6 focused on the reading process and reading comprehension rubrics. The student text annotations and responses to the reading process questions were scored by two different raters. The final scores on the Reading Process and Text Summary represent the average score across the two scores.

Reliability of Scores

Reliability of Writing Scores

In order to check for reliability of the ILA Writing Content and Language scores, we conducted a generalizability study (G-study) with all trained raters on 20 randomly selected student assessments for the two rubrics evaluating the essay. G-study provides systematic examination of the different sources of error in the scores and their relative importance in the score variability. The results from the G-studies suggest that the rater reliability on the essay dimensions were generally high.

As shown in Table 5, the vast majority of the variation in students' scores, in both content and language was due to variation in the student papers themselves, 68.0% and 64.0% respectively. Variation in scores due to rater inconsistency was low for both rubrics: differences among the raters accounted for only 4% of the total variability in the Content scores, 2% of the total variability in the Language scores. Based on our calculated g-coefficients, the reliabilities across the raters were relatively high for both content (0.96) and language (0.95).

Table 5

G-study Summary Table

Component	Content	Language
Var(rater)	0.03 (4.0%)	0.02 (2.0%)
Var(paper)	0.51 (68.0%)	0.55 (64.0%)
Var(paper*rater)	0.21 (28.0%)	0.29 (34.0%)
G-coefficient	0.96	0.95

Reliability of Reading Scores

For the Reading Process scores, we examined the reliability across the two raters using the intra-class correlation. The Intra-class Correlation (ICC) assesses rater reliability by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects. The range of the ICC is between 0.0 and 1.0. The ICC will be high when there is little variation between the scores given to each paper by the two raters. For Reading Process, we calculated ICC. The ICC for 623 papers scored by two raters is 0.56 for Reading Process scores. This suggests that the raters were only moderately consistent in their ratings across the papers.

To examine the rater reliability for the ratings on the Text Summary, we calculated ICC. The ICC for 614 papers scored by two raters is 0.80. This suggests that the raters were only fairly consistent in their ratings across the papers.

Reliability of MC Items on the Content Understanding and Reading Comprehension

The reliability of the items for the Content and Reading Comprehension was evaluated using an internal consistency measure. Internal consistency measures are indicators of how well the items for each construct relate to each other. For a good measure of internal consistency, the alpha coefficient should be fairly high (e.g., > 0.80). The Cronbach's alpha for the five items on the Content Understanding was 0.51. The reliability for the Reading Comprehension was slightly higher than the Content items with the alpha at 0.60. Both of these sections had only moderately high internal consistency among the items.

Correspondingly, the knowledge of biology contributes to scientific literacy and helps to understand the world in which we live [5, 6]. However, the studentsâ€™ learning performance and achievement in biology have been poor over recent years [7]. The poor performance and laboratory proficiency in biology are influenced by not having expository approaches that stand up to challenge the objectives of biology education [5, 8â€“10].Â 3.1. The Impact of CL on the Studentsâ€™ Academic Achievement in Biology Subject.Â in the studentsâ€™ practical skill performance could be attributed to the lack of laboratory equipment and technicians, teachersâ€™ commitment, and studentsâ€™ initiation to be engaged in laboratory work.