

Science Achievement of English Language Learners in Urban Elementary Schools: Results of a First-Year Professional Development Intervention

Okhee Lee,¹ Jaime Maerten-Rivera,¹ Randall D. Penfield,¹ Kathryn LeRoy,² Walter G. Secada³

¹*School of Education, University of Miami, Coral Gables, Florida 33146*

²*Miami-Dade County Public Schools, Elementary Curriculum and Instruction,
Miami, Florida, 33132*

³*School of Education, University of Miami, Coral Gables, Florida 33146*

Received 11 January 2006; Accepted 20 February 2007

Abstract: This study is part of a 5-year professional development intervention aimed at improving science and literacy achievement of English language learners (or ELL students) in urban elementary schools within an environment increasingly driven by high-stakes testing and accountability. Specifically, the study examined science achievement at the end of the first-year implementation of the professional development intervention that consisted of curriculum units and teacher workshops. The study involved 1,134 third-grade students at seven treatment schools and 966 third-grade students at eight comparison schools. The results led to three main findings. First, treatment students displayed a statistically significant increase in science achievement. Second, there was no statistically significant difference in achievement gains between students at English to Speakers of Other Language (ESOL) levels 1 to 4 and students who had exited from ESOL or never been in ESOL. Similarly, there was no significant difference in achievement gains between students who had been retained on the basis of statewide reading test scores and students who had never been retained. Third, treatment students showed a higher score on a statewide mathematics test, particularly on the measurement strand emphasized in the intervention, than comparison students. The results indicate that through our professional development intervention, ELL students and others in the intervention learned to think and reason scientifically while also performing well on high-stakes testing. © 2007 Wiley Periodicals, Inc. *J Res Sci Teach* 45: 31–52, 2008

Keywords: general science; diversity; professional development; elementary; statistics/multivariate

Education reform in general, and the *No Child Left Behind* [NCLB] Act of 2001 in particular, require that *all* students achieve high academic standards in core subject areas. Teachers working with students from diverse languages and cultures—in short, most of today’s teachers—face the

Contract grant sponsor: National Science Foundation; Contract grant number: ESI–0353331.

Correspondence to: O. Lee; E-mail: olee@miami.edu

DOI 10.1002/tea.20209

Published online 14 June 2007 in Wiley InterScience (www.interscience.wiley.com).

challenge of making academic content and process accessible and meaningful for their students. Ideally, content area instruction should provide a meaningful learning environment for English language and literacy development, while improving English skills should provide the medium for understanding academic content (Casteel & Isom, 1994; Lee & Fradd, 1998). In reality, however, English language learners (ELL) students frequently confront the demands of academic learning through a yet-unmastered language *without* the instructional support they need. Thus, professional development and classroom practices should integrate academic disciplines with English language and literacy with ELL students. Such integration is urgent, given the policy context of standards-based instruction, high-stakes testing, and accountability facing today's schools.

Our research addresses low science achievement of ELL students in the context of the impending high-stakes testing and accountability policy in science that will begin to occur nationally in the 2007–2008 school year under the *NCLB* Act and in the 2006–2007 school year in the state in which the research takes place. The impact of this policy change is greater with ELL students, as more states are shifting from bilingual education to “English only” policies that fail to consider students' proficiencies in the home language as relevant to academic achievement. Over the course of its 5-year period using longitudinal designs, the research implements a professional development intervention (consisting of curriculum units and teacher workshops) that is aimed at improving science and literacy achievement of ELL students in urban elementary schools within the policy context increasingly driven by high-stakes testing and accountability across content areas. The research tests two conventional wisdoms: (a) can ELL students learn academic subjects, such as science, while also developing English proficiency? and (b) can ELL students, who learn to think and reason scientifically, also perform well on high-stakes testing? The research involves teachers from grades 3 through 5 and their students at 15 elementary schools in a large urban school district. All the schools enroll high proportions of ELL students and students from low socioeconomic status (SES) backgrounds, and have traditionally performed poorly according to the state's accountability plan.

As part of the larger research project, this study tested the second conventional wisdom by examining third-grade students' science achievement after the first-year implementation of our professional development intervention. (To test the first conventional wisdom, the larger project administered a writing prompt as a measure of literacy development in the beginning and at the end of the school year. These results are reported elsewhere.) Specifically, the study addressed the following research questions:

1. Did the students in the treatment group display gains in science achievement (as measured by a project-developed science test) from pretest to posttest of the first-year of implementation?
2. Did the gaps in science achievement between students at different levels of English proficiency and literacy change from pretest to posttest in the treatment group?
3. Did the treatment and comparison groups perform differently on the high-stakes statewide mathematics test, particularly the measurement strand emphasized in the intervention?

Literature Review

Many teachers, especially at the elementary level, are not adequately prepared in subject areas, such as science, or in subject-specific teaching strategies (Kennedy, 1998). Additionally, teachers are not sufficiently prepared to meet the learning needs of ELL students (National Center for Education Statistics, 1999). Thus, it is essential to provide elementary teachers with professional development opportunities in teaching science and English proficiency simultaneously with ELL students in urban schools.

Effective Science Instruction with ELL Students

With ELL students, English language and literacy development is integral to content area instruction, such as science (Amaral, Garrison, & Klentschy, 2002; Hampton & Rodriguez, 2001; Lee & Fradd, 1998; Stoddart, Pinal, Latzke, & Canaday, 2002). English proficiency involves knowledge and effective use of the conventions of literacy, such as vocabulary, syntax, spelling, and punctuation, in social and academic settings. In content areas, proficiency includes knowledge of various subregisters representing specific disciplines. In addition, science employs nontechnical terms that have meanings unique to scientific contexts (e.g., matter, force, energy, space). Language functions (e.g., describing, hypothesizing, explaining, predicting, and reflecting) develop simultaneously with science inquiry and process skills (e.g., observing, describing, explaining, predicting, estimating, representing, inferring) (Casteel & Isom, 1994). Furthermore, students learn science through thinking and reasoning as members of a science learning community.

Research on science instruction with ELL students focuses on hands-on, inquiry-based science instruction to promote science learning and English proficiency simultaneously (Amaral et al., 2002; Casteel & Isom, 1994; Lee, Deaktor, Hart, Cuevas, & Enders, 2005). First, hands-on activities are less dependent on formal mastery of the language of instruction, thus reducing the linguistic burden on ELL students. Second, hands-on activities through collaborative inquiry support language acquisition in the context of authentic communication about science knowledge. Third, inquiry-based science promotes students' communication of their understanding in a variety of formats, including written, oral, gestural, and graphic. Finally, by engaging in the multiple components of science inquiry, ELL students develop their English grammar and vocabulary as well as their familiarity with scientific genres of speaking and writing.

Science instruction for ELL students should be conceptualized and implemented within the constraints of urban schools where these students tend to be concentrated. Many elementary classrooms lack appropriate science instructional materials and supplies, a state of affairs often exacerbated by a more generalized lack of resources and funding in urban schools (Knapp & Plecki, 2001; Spillane, Diamond, Walker, Halverson, & Jita, 2001). Additionally, instructional time for science in low-performing urban elementary schools is often limited and tightly regulated due to the urgency of developing basic literacy and numeracy in students with limited skills and those learning English as a new language. Furthermore, teachers wishing to pursue inquiry-based science instruction in urban schools face added challenges in the context of high-stakes testing and accountability, as sanctions against poor academic performance are disproportionately leveled against urban schools (Settlage & Meadows, 2002).

Professional Development for Effective Science Instruction with ELL Students

To provide effective science instruction, teachers need opportunities to develop their own deep and complex understandings of science concepts and recognize how students' misconceptions cause learning difficulties (Kennedy, 1998). Teachers also need to engage in science inquiry themselves to be able to foster student initiative in inquiry (National Research Council, 2000). Additionally, teachers need to learn how to enable students to share and negotiate ideas and construct collective meanings about science (Lemke, 1990).

In addition to ensuring that ELL students acquire the communicative language functions used for social language, teachers should create classroom environments that promote the development of academic language (Wong-Fillmore & Snow, 2002). Also, teachers should view language from a human development perspective and formulate developmentally appropriate expectations about

language comprehension and production over the course of learning English. Finally, teachers should apply this knowledge to the teaching of academic content areas. The amalgamation of these three knowledge sources should result in teaching practices that engage students of all levels of English proficiency in academic language learning, offer multiple points of entry for students of differing levels of English proficiency, and provide multiple modes for students to display their learning.

Research on professional development interventions to promote science achievement for ELL students is limited, but has begun to emerge in recent years (see the review of literature by Lee, 2005). Several studies examined the impact of professional development on ELL students' science achievement. For example, Amaral et al. (2002) examined professional development in promoting science with predominantly Spanish-speaking elementary students as part of a district-wide local systemic reform initiative. Teachers received professional development, in-classroom professional support from resource teachers, and complete materials and supplies for all the science units. Students in the district participated in kit- and inquiry-based science instruction that included the use of science notebooks. All students were assessed using the Stanford Achievement Test that served as the statewide science assessment. Results indicated that science achievement increased in direct relation to the number of years they participated in the program. English proficient students performed significantly better than limited English proficient students.

Lee and colleagues implemented a professional development intervention aimed at promoting science achievement for culturally and linguistically diverse elementary students. Through the provision of curricular materials and teacher workshops, the intervention focused on integrating English language and literacy development as part of science instruction with ELL students. On paper-and-pencil science tests administered to all third- and fourth-grade students, ELL students showed statistically significant gains, and achievement gaps narrowed (Lee et al., 2005). Additionally, on performance assessment tasks with a smaller sample, ELL students demonstrated enhanced abilities to conduct science inquiry, and achievement gaps narrowed (Cuevas, Lee, Hart, & Deaktor, 2005).

Several studies have examined the impact of curriculum materials on ELL students' science achievement. For example, Hampton and Rodriguez (2001) implemented a hands-on, inquiry science curriculum (i.e., the Full Option Science Series, FOSS) with Spanish-speaking elementary students who were developing a second language (English) along with their first language (Spanish). On a written assessment containing three inquiry items and three open-ended response items about foods and nutrition, correct performance ranged from 33% to 51% across the six items. There was no significant difference between students who chose to respond in Spanish and those who chose to respond in English.

Fradd, Lee, Sutman, and Saxton (2002) developed and tested curriculum materials that integrated science inquiry, home language and culture, and English language and literacy of ELL students. Elementary students from different ethnolinguistic backgrounds, including Spanish, Haitian Creole, and monolingual English-speaking students of White and African-American descent, completed the Matter and Weather units. At the beginning and end of each unit, students completed a paper-and-pencil test containing multiple-choice, short answer, and extended response items. Students from all ethnolinguistic groups showed statistically significant achievement gains in science knowledge and inquiry.

In summary, this literature provides insights for developing effective professional development interventions to promote science achievement of ELL students. Building on this emerging literature, our intervention is focused on professional development of urban elementary school teachers through the provision of curriculum units and teacher workshops within the policy context increasingly driven by high-stakes testing and accountability. This study

specifically examined the impact of the first-year of the intervention on science achievement of ELL students—did they learn to think and reason scientifically, while also performing well on high-stakes testing?

Research Setting and Participants

State in which Research was Conducted

Public schools in the state are assigned a letter grade (A, B, C, D, or F) based on a formula from the state's school accountability plan. Currently, the school grade is based on student performance in reading and mathematics from grades 3 through 5 and writing at grade 4. Starting in 2002–2003, third-grade students were retained if they received a Level 1 achievement score on the statewide reading assessment. A statewide science assessment was administered at the fifth-grade level beginning in the 2002–2003 school year, but does not yet count toward the school grades on which accountability is based until the 2006–2007 school year.

Under the “English-only” policy regarding ELL students, the state implements English to Speakers of Other Languages (ESOL) programs focusing on the acquisition of the English language with little attention to the maintenance or development of the home language. Currently, ELL students are held accountable for school and individual performance on statewide assessments 2 years after their school enrollment.

School District

The research was conducted in a large urban school district in the southeast U.S. with a student population displaying a high level of linguistic and cultural diversity. During the 2004–2005 school year, the ethnic makeup of the student population in the school district was 60% Hispanic, 28% Black, 10% White non-Hispanic, and 2% Asian or Native American. Across the school district, 72% of elementary students participated in free or reduced price lunch programs, and 24% were designated as limited English proficient (LEP), which is the term used by the state to designate ELL students in ESOL programs. These three terms are comparable; the terms ELL and ESOL students are used interchangeably in this article, but the term LEP students is not used.

Schools

In late May 2004, elementary schools were selected based on three criteria: (a) percentage of ELL students (predominantly Spanish or Haitian Creole-speaking students) above the district average at the elementary school level, (b) percentage of students on free and reduced price lunch programs above the district average at the elementary school level, and (c) school grades of primarily C or D according to the state's accountability plan since its inception in the 1998–1999 school year.

Of the 206 elementary schools in the district, 33 met these criteria. Of these schools, 17 volunteered to participate starting in the fall of 2004. Based on a set of criteria, eight schools were assigned to the treatment group and nine schools to the comparison group. Shortly after the project commenced, one treatment and one comparison school each withdrew. Thus, the research during the first year involved 15 elementary schools, including seven in the treatment group and eight in the comparison group.

For our school-wide initiative, all third- through fifth-grade teachers in each treatment school will eventually participate in the research. The treatment group schools engage in the intervention for 3 years and continue a 1-year follow-up without the intervention to test sustainability. A subset of the comparison schools will engage in the intervention during the remaining 3 years of the project to test replicability.

Teachers and Students

During the first year of the research (2004–2005), only third-grade teachers in the seven treatment schools participated in the intervention, for a total of 42 teachers among 41 classrooms (one classroom had two teachers). In terms of demographic backgrounds, 38 teachers were female; 4 male. Of these teachers, 17 identified themselves as Black, 13 as Hispanic, 7 as White non-Hispanic, 3 as Haitian, and 2 as Asian. In addition, 32 teachers reported that English was their native language, 8 Spanish, and 2 Haitian Creole. In terms of professional backgrounds, 1 teacher reported having a specialist degree (beyond master's degree), 14 had master's degrees, and 27 had bachelor's degrees. Their teaching experience ranged from 1 to 33 years, with an average of 10 years. They had been teaching at their current schools for an average of 6 years.

Table 1 provides the demographic makeup of the third-grade students in the treatment and comparison schools. The comparison schools enrolled proportionately fewer Hispanic students, more Black students, more ESOL students, and more retained students than the treatment schools.

Professional Development Intervention

The professional development intervention was comprised of (a) curriculum units, including student booklets, teachers' guides, and science supplies, and (b) teacher workshops throughout the school year. The intervention's potential impact on student achievement was mediated by teachers' classroom practices with their students.

Table 1
Third grade student demographics (%)

Variable	Demographic Groups	Treatment Schools (<i>N</i> = 1,134)	Comparison Schools (<i>N</i> = 959)
Gender	Male	50%	49%
	Female	50	51
Ethnicity	Hispanic	52	47
	Black (including Haitian and Caribbean immigrants)	44	50
	White non-Hispanic	3	2
	Other	1	1
ESE	Exceptional students (not including gifted students)	12	12
ESOL	ESOL levels 1 through 4	15	22
	Exited from ESOL within 2 years	38	31
	Exited from ESOL over 2 years or never in ESOL	47	47
Retention	Retained at least once	19	25

Note: Demographic information was missing for 112 treatment group students and 7 comparison group students. The number of missing data was higher for the treatment group because the data for the treatment group were collected at multiple time points and then matched with school district data, whereas the data for the comparison group were collected at one point from the school district.

ESOL, English to speakers of other languages; ESE, exceptional student education.

Curriculum Units for Teachers and Students

Throughout the project period, a series of nine curriculum units will be developed that constitute the entire science curriculum for grades 3 through 5 as mandated by the state science content standards and also recommended by the *National Science Education Standards* (National Research Council, 1996). The three units for grade 3 (Measurement, States of Matter, and Water Cycle and Weather) were developed and rigorously tested in our previous research, and the units for grades 4 and 5 are currently under development. The curriculum units emphasize science topics that are assessed annually, along with those topics that are assessed once every 3 years, according to the state's accountability plan. The third-grade students participating in the first year of our intervention will be the first cohort to take the high-stakes science test to be administered at the fifth-grade level.

The materials development team consists of scientists, science educators, bilingual/ESOL educators, mathematics educators, and district administrators in science education. A critical component was the involvement of elementary teachers from our previous research. Based on their experience and understanding of the overall goals of the research, they provided insights about linguistic and cultural practices of diverse student groups, the appropriateness of science content and inquiry skills for elementary students, the feasibility of implementing the intervention in elementary classrooms, and the relevance of the curriculum materials for high-stakes statewide testing in reading, writing, and mathematics (science is not part of high-stakes testing and accountability until 2006–2007). As we continue curriculum development efforts, we will revise and refine the units based on teacher feedback, classroom observations, and student assessment results in our current research. For example, our ongoing efforts to incorporate teachers' feedback to our curriculum units in particular and our intervention more broadly are reported elsewhere (Lee, LeRoy, Thornton, Adamson, Maerten-Rivera, & Lewis, in press).

The teachers' guide for each unit begins with an explanation of: (a) how to promote students' science inquiry and understanding of key science concepts and "big ideas" (patterns of change, systems, models, and relationships) to explain natural phenomena, (b) how to incorporate English language and literacy development as part of science instruction, and (c) how to incorporate mathematics to support science instruction. For each lesson, the teachers' guide includes (a) specific correlations to state content standards in science, language arts, and mathematics; (b) key vocabulary terms in English, Spanish, and Haitian Creole; (c) glossary of science vocabulary; (d) a list of materials for each hands-on activity; (e) transparencies of pictures, drawings, tables, graphs, and charts; and (f) specific teaching suggestions to support student learning. Additionally, the teachers' guide offers suggestions for writing prompts, field trips, and trade books or literature related to the science topics. Below, key elements of the units in terms of science, English language and literacy, and mathematics are described.

Science. The units were developed to promote student initiative and responsibility in conducting inquiry, as teachers gradually reduce their level of guidance. According to the *National Science Education Standards* (National Research Council, 1996), "Students will engage in selected aspects of inquiry as they learn the scientific way of knowing the natural world, but they also should develop the capacity to conduct complete inquiries" (p. 23). The units are designed to move progressively along the *teacher-explicit to student-initiated continuum* to promote science inquiry. Providing more structure to earlier lessons within each unit, while later lessons are more open-ended, encourages student initiative and exploration. The level of complexity of science concepts and degree of science inquiry required from students also increase as students move through the units. Far from being scripted, the units offer a great deal of flexibility and

openness. Each lesson ends with science inquiry practices—from a list of questions provided, students select those questions that would be suitable for science inquiry; raise their own questions related to the science inquiry activity in the lesson; and evaluate these questions based on a set of criteria that describe what constitutes good science inquiry. At the end of each unit, students engage in an inquiry extension by conducting open inquiries based on their own questions related to the content within the unit. Within the context of science inquiry, student booklets emphasize key science concepts and big ideas. Following inquiry activities, each lesson provides science background information that explains the question under investigation and related natural phenomena. The lesson also highlights common misconceptions and potential learning difficulties that students may have.

Teachers' guides provide content-specific teaching strategies. They offer suggestions on how teachers may provide different levels of guidance and scaffolding depending on students' prior experience with different science topics and the demands of specific science tasks. They also offer suggestions about how to set up and implement hands-on activities, along with cautions about what may go wrong and how to respond to such situations. Additionally, they provide science background information and explanations for the questions posed in the student booklets, with particular emphasis on students' common misconceptions and learning difficulties. Furthermore, they offer suggestions for extension activities, assessment activities, and homework assignments.

English Language and Literacy. Student booklets highlight activities or strategies to foster reading and writing as part of science instruction. For example, the booklets use specific comprehension questions about inquiry activities, strategies to enhance comprehension of science information in expository text at the end of each lesson, and various language functions (e.g., describing, explaining, reporting, drawing conclusions) in the context of science inquiry. Teachers' guides also provide suggestions to promote literacy development. For example, students engage in authentic communication through the use of hands-on tasks, narrative vignettes, and expository texts related to everyday experiences. Students write expository paragraphs describing the scientific process under investigation, explanations, and conclusions of science experiments conducted in class, or responses to the writing prompts provided as supplementary materials. Trade books or literature related to the science concepts under investigation are incorporated.

In addition to general literacy development in English for all students, the units address the needs of ELL students by providing explicit guidance to promote their English proficiency. For example, science terms in Spanish and Haitian Creole are provided to support communication and comprehension. Language load for students at varying levels of English proficiency is increasingly more demanding from grades 3 through 5. The units introduce key vocabulary in the beginning and encourage students to practice the vocabulary in a variety of settings to enhance their understanding throughout the lesson and over the course of the unit. Additionally, the units use multiple modes of communication and representation (verbal, gestural, written, graphic) to enhance students' understanding of science. Teachers' guides also emphasize the importance of linguistic scaffolding to promote ELL students' comprehension and understanding of science. For example, extensive graphic materials are included in transparencies (e.g., graphic organizers, Venn Diagrams, pictures of measurement instruments, drawings of experimental setups, data tables, graphs, charts). Teachers are encouraged to engage students in a variety of group formations, so that students learn to communicate independently, in small groups, and with the whole class.

Mathematics. Although the focus of the curriculum is science and literacy, the units also consider mathematics (National Council of Teachers of Mathematics, 2000) as an important supporting area for science learning. Particularly, our intervention focuses on measurement as basic skills and abilities for engaging in science inquiry. The topic of measurement is tested as one of the five strands of mathematics content standards on the statewide mathematics assessment.

The intervention begins with a comprehensive unit on measurement that is aligned with the state-mandated benchmarks within the measurement strand for third grade. First, students learn measurement skills, including metric and traditional systems of measurement, units of measurement, increments in instruments, and calibration of instruments. They compare and contrast the metric and traditional systems and understand the appropriate use of units for each system. Second, students learn conventions for measurement instruments—ruler and measuring tape for length; kitchen scale, bathroom scale, pan balance, and triple beam balance for weight; measuring cup and graduated cylinder for volume; thermometer for temperature; and clock and stop watch for time. They develop estimation strategies as well as skills for precise measurements. Finally, students learn conventions for recording and displaying measurement data using charts, tables, and graphs.

As students engage in science inquiry in the Matter and Water Cycle/Weather units, they employ measurement skills and concepts they have learned from the Measurement unit. Over the course of the year, students become precise and accurate in taking measurements, identifying patterns and anomalies in data, using multiple representational formats for data displays, and reasoning quantitatively.

Teacher Workshops

During the first year of the project, third-grade teachers in the treatment group attended 5 full-day workshops on regular school days. Project personnel with expertise in science education, ESOL, mathematics education, and linguistic and cultural issues in education designed and conducted the workshops. Teachers were actively involved as they shared questions, suggestions, and examples of their own practices and beliefs. Teachers also shared their thoughts about similarities and differences in teaching and learning environments among the participating schools.

The first workshop was organized around describing the purpose of the project, obtaining teachers' consent, conducting data collection activities, and introducing the first few lessons of the Measurement unit. The second workshop was organized around the Measurement unit, the third workshop around the Matter unit, the fourth workshop around the Water Cycle and Weather unit, and the final workshop around data collection activities including teachers' reflection and feedback. Each workshop was conducted twice, with half of the teachers attending the first session and the other half attending the second; this was done to reduce the number of participants at each workshop and to eliminate concerns from large schools where all of the third-grade teachers would have been out on the same day.

Science. As part of our first-year efforts for professional development in a longitudinal design, the workshops focused on familiarizing teachers with the science content, hands-on activities, common student misconceptions, and potential learning difficulties in each lesson. The third workshop on the Matter unit introduced and emphasized inquiry-based science. Project personnel and teachers discussed what science inquiry involves (National Research Council, 1996, 2000) and how teachers promote student initiative in conducting inquiry as teachers

gradually reduce their level of guidance. They also discussed the notion of the teacher-explicit to student-initiated continuum in providing instructional scaffolding to promote science inquiry. Effective inquiry instruction requires a balance of teacher guidance and student initiative, as teachers make the decisions about when and how to foster student responsibility. Teachers discussed how to move away from teacher-explicit instruction and encourage students to take the initiative and assume responsibility for their own learning. Based on this discussion, teachers worked in small groups on lessons from the Matter unit. Then, using inquiry tasks as examples, project personnel demonstrated how to structure science instruction around inquiry activities. Given “practice” inquiry tasks, small groups of teachers came up with a variety of experimental designs, procedures for gathering data, multiple ways of displaying the data, and conclusions based on hypothetical evidence. Each group of teachers presented their work to the entire group and discussed various ways of conducting science inquiry. The emphasis on how to promote more open-ended and student-centered inquiry continued with the Water Cycle and Weather unit at the fourth workshop.

Scientific reasoning was emphasized throughout the workshops. The focus was to identify students’ cultural and linguistic experiences from their home environments that could serve as intellectual resources for learning school science, as well as their difficulties with science concepts and inquiry. During the second workshop on the Measurement unit, teachers brought their own students’ work samples from the Measurement student booklet and discussed student reasoning of measurement concepts and skills. At the end of the second workshop, teachers completed state-released practice items on a statewide science test. Teachers’ reasoning about their test responses was one focus of the third workshop. At the fourth workshop on the Matter unit project personnel presented our previous research on students’ reasoning about designing an experiment to test the effect of surface areas on the rate of evaporation. Using segments of students’ interview transcripts, teachers analyzed students’ capabilities and difficulties in designing the experiment. At the final workshop, project personnel presented major patterns in students’ reasoning about various measurement concepts and skills from our current research. The presentation highlighted the “funds of knowledge” about measurement in students’ home environments that could serve as building blocks for learning school science. Based on this presentation, teachers discussed how their students made sense of home experiences and how they could use these home connections to promote student learning. For example, in the Haitian community where this research is being conducted, women often work as seamstresses. A teacher shared the example by one of her Haitian students whose grandmother measured the length of fabric by using her open arm as close approximation of 1 yard and being slightly short of 1 meter. This example supported the real world connection to estimation and units of measurement.

The state science content standards were emphasized as a backbone of the workshops. At the first workshop, project personnel described how the curriculum units from grades 3 through 5 would align to the state science content standards. At the second workshop, teachers became familiar with the benchmark clarification of those standards that are being assessed at grade 5. Teachers were introduced to the state-defined content clusters including those benchmarks that are annually assessed as well as those assessed every 3 years. Teachers also became aware of assessment item formats and probable impacts of high-stakes science test results on school grades according to the state’s accountability system. For each curriculum unit at each workshop, project personnel demonstrated how the unit corresponded to specific science benchmarks. Especially, project personnel helped teachers recognize how students’ science inquiry and reasoning abilities could enhance performance on statewide science assessment.

English Language and Literacy. The workshops focused on incorporating English language and literacy development into specific science lessons. At the second workshop, using examples in the Measurement unit, project personnel described various strategies for developing students' reading and writing skills as part of science lessons. Project personnel also described how to provide linguistic scaffolding for ELL students. The discussion focused on how teachers: (a) adjust the level and mode of their communication (verbal, gestural, written, graphic) to enhance students' understanding of science; (b) recognize the diversity of students' levels of language proficiency and adjust the language load required for their participation; (c) use language that matches students' levels of communicative competence in length, complexity, and abstraction; and (d) communicate at or slightly above students' levels of communicative competence. Based on this presentation, teachers engaged in a jigsaw activity regarding how to incorporate ESOL strategies into science lessons. At the fourth workshop, teachers worked in small groups to incorporate ESOL strategies in selected lessons from the Water Cycle and Weather unit. As a culminating activity, teachers made group presentations followed by whole group discussion.

Mathematics. Although the intervention's focus is on science and literacy, the workshops also emphasized the role that mathematics plays in science inquiry. At the first and second workshops on the Measurement unit, teachers engaged in measurement tasks, including length, weight, volume, temperature, and time. This allowed the teachers to discuss various aspects of measurement, such as accuracy in measurement, error in measurement, increments in instruments, calibration of instruments, and estimation versus precise measurement in different settings. At the third and fourth workshops on the Matter and Water Cycle/Weather units, project personnel emphasized how mathematics concepts and skills are fundamental to engaging in science inquiry, including measurement, recording and display of data using multiple representational formats (e.g., graphs, charts, tables, drawings), and patterns and anomalies in the data. In the context of conducting experiments, teachers discussed measurement issues, such as accuracy in measurement, multiple trials, and analyses and interpretations of data. These mathematical concepts and skills through all aspects of science inquiry strengthen both mathematics and science learning.

Mediating Classroom Practices

Teachers were provided with complete class sets of materials, including teachers' guides, copies of student booklets, science supplies designed for six small groups per class, and trade books related to the science topics. Classroom science instruction took place on average 2 or 3 hours a week, which was consistent with the school district guideline mandating 150 minutes a week for science instruction. Most teachers completed instruction of the three third-grade units by the end of the school year.

Classroom observations were employed twice during the school year (once in the fall of 2004 and once in the spring of 2005 for a total of close to 90 classroom observations) to examine classroom practices with regard to scientific understanding, scientific inquiry, teachers' knowledge of science content, and teacher support of English language development (for details, see Lee, Lewis, Adamson, Maerten-Rivera, & Secada, 2006). Based on the means and modes of classroom observation ratings, common patterns of classroom practices were identified, as briefly described below.

First, students' scientific understanding was uneven in most observed lessons. Students grasped deep understanding of some scientific concepts and ideas, but superficial understanding of

others. Second, classroom practices to promote students' scientific inquiry were noticeably different between fall and spring. During the Measurement unit in fall, teachers taught measurement as basic skills or tool use, and students rarely engaged in higher level reasoning associated with estimation, prediction, posing a question, or finding a solution. During the Water Cycle/Weather unit in spring, students conducted scientific inquiry within the bounds of scripted lessons, as they primarily received and performed routine procedures for the inquiry. Third, teachers' knowledge of science content was generally accurate within the bounds of the lesson content provided in the student booklets and the teachers' guides, but most teachers did not demonstrate deeper knowledge beyond the bounds of the lesson content. Finally, teachers communicated at the appropriate level or mode of language and sometimes used language support strategies (e.g., multiple modes of representation, use of language in multiple settings, and use of ELL students' home language as needed). Overall, all four areas of classroom practices were generally within the bounds supported by the intervention. However, such practices fell short of the goal of reform-oriented practices in science instruction (National Research Council, 1996, 2000) or English language and literacy development in content area instruction (Teachers of English to Speakers of Other Languages, 1997).

Research Procedures

Tests

The project team developed a science test to assess students' knowledge of key science concepts and big ideas of patterns, systems, models, and relationships for the science topics of the curriculum units during the school year. The test also measured students' understanding of science inquiry using relatively structured inquiry tasks (similar to NAEP performance tasks) in which students construct graphs and tables using the data provided, offer an explanation for the data, and draw a conclusion. The test was developed in English to maintain continuity between the language of instruction and the language of assessment. Item formats included multiple-choice, short answer, and extended responses. The test consisted of a total of 10 items, and some of them had subcomponents. It contained six project-developed items that have an item format consistent with the state science assessment, two public-release NAEP items from the 1996 and 2000 Grades 4 and 8 Science subject tests, and two public-release TIMSS items from the Mathematics and Science release pool for grades 3/4 and 7/8. Teachers indicated that the length of the tests (6 pages) was appropriate for third-grade students.

Although the project-developed science test measured students' understanding of science concepts and inquiry within the treatment group, a statewide mathematics test was used to measure students' performance on high-stakes testing in both the treatment and comparison groups. The state test in mathematics is given at grades 3 through 10 and assesses five strands: (1) number sense, concepts, and operations; (2) measurement; (3) geometry and spatial sense; (4) algebraic thinking; and (5) data analysis and probability. Students are required to complete primarily multiple-choice items and a few short answer and extended response items with cognitive demand measured by item difficulty and cognitive complexity. Our intervention directly addressed the measurement strand of the statewide mathematics test.

Data Collection and Coding

The project-developed science test was administered to the third-grade students in the treatment group at the beginning and end of science instruction over the school year. Teachers

followed standard procedures for test accommodations with ELL students and students with limited literacy. Teachers read the items for students with reading difficulties. Teachers who spoke the home language of their ELL students translated the items in the students' home language, while teachers who did not speak their students' home language solicited help from another colleague or used the translated science terms in Spanish and Haitian Creole provided in the teachers' guide for each unit. ELL students were allowed to write their answers in either English or in their home language. There was no time limit in completing the test.

The maximum points for each specific item or item subcomponents ranged from 1 to 3, depending on the level of cognitive or conceptual difficulties, with a score of 0 used to indicate irrelevant or no response. For items with two or three points, a scoring rubric was developed to assess the conceptual accuracy of responses, completeness of responses, and use of science terms. Students were awarded partial credit on items as warranted. The total possible score on the test was 24 points.

After training by a consultant who had been involved in the development of the scoring rubric from our previous research, a group of seven district science staff scored student responses on pretests and posttests. The interrater agreement among the coders was established at 90%. Throughout the scoring process, coders consulted with one another to ensure agreements on scores.

Internal consistency reliability estimates of the test scores were .60 for the pretest and .71 for the posttest. The reliability estimate for the posttest was within an acceptable range, whereas the reliability estimate for the pretest tended to be lower than generally acceptable.

Data Analysis

The analyses were based on two measures of student achievement. The first involved scores on a science test obtained in the fall (pretest) and spring (posttest) for all students in the treatment group. Each student received a gain score in science achievement, which was computed by subtracting the pretest score from the posttest score. The second measure involved scores on five strands of the statewide mathematics test obtained for all students in the treatment and comparison groups.

There were 1,027 students who took the science pretest and 925 students who took the posttest in the treatment group. Analyses were conducted with 818 students who took both the pretest and posttest. The number of students omitted from the analyses was 316, or 28% of the 1,134 students in the entire student sample. The missing data resulted primarily from two teachers failing to complete posttests as well as general student mobility. These kinds of losses in the sample are common in field-based research, especially in urban schools with high student mobility.

For the statewide mathematics test that was administered in mid-March, the test scores for both the treatment and comparison group students were obtained from the school district database. If a student from the treatment group took either the pretest or the posttest, the student was included in the analysis; however, we were unable to obtain the test scores of some of these students. In the comparison group, all students enrolled in third grade were included. The analyses were conducted with 942 students in the treatment group and 966 students in the comparison group.

Independent (or explanatory) variables included in the analyses consisted of gender, ethnicity, exceptional student education (ESE) classification, ESOL classification, retention classification, and treatment versus comparison group. Ethnicity was recorded as Hispanic, Black, White, or other. Because 95% of the sample in the treatment and comparison groups was either Black or Hispanic, comparisons among ethnic groups are largely reflective of a Black versus Hispanic

comparison. ESE classification was recorded as either ESE or non-ESE, where non-ESE also included students classified as gifted. The information about the ESE classification was obtained at the time when statewide assessments were administered in mid-March. ESOL classification was recorded using the categories ESOL levels 1 to 4, in contrast to students who had exited from ESOL programs or had never been in ESOL programs. The information about the ESOL classification was obtained along with the information about the ESE classification in mid-March. Retention classification was recoded into a dummy variable that contrasted students who had never been retained with those who had been retained at least once. Finally, the treatment variable was applicable only to the analyses using the strand scores of the statewide mathematics test because these scores were obtained for both the treatment and comparison groups, whereas the science test was administered only to the treatment group. The SES variable was not considered, because all the participating schools had close to or over 90% of students in free or reduced price lunch programs.

The ESOL classification presents challenges conceptually and methodologically. ESOL programs typically include students at ESOL levels 1 through 4. Once students are deemed English proficient, they exit from ESOL programs but remain as ESOL level 5 for 2 years. These students still require instructional support to learn the academic register of English in subject areas, including science. In this study, the treatment schools had 15% of students at ESOL levels 1 through 4 and 38% at ESOL level 5, and the comparison schools had 22% of students at ESOL levels 1 through 4 and 31% at ESOL level 5 (see Table 1). Although it would be desirable to analyze the ESOL variable into three groups (i.e., ESOL levels 1 through 4, ESOL level 5, and students who exited over 2 years and non-ESOL students), the sparseness of the data for ESOL levels 1 through 4 within each classroom posed a threat to the stability of parameter estimates in the hierarchical linear modeling (HLM) analyses. Given that ESOL level 5 students tend to perform higher than ESOL levels 1 through 4 (according to the analyses in this study; also see Cuevas et al., 2005; Lee et al., 2005), it was deemed most appropriate to collapse ESOL levels 1 to 4 into one category, and all other students into a second category. Although we acknowledge that this dichotomization is somewhat crude, it provided the most effective mechanism for (a) guarding against the sparseness of the ESOL levels 1 to 4 observed in the current sample, (b) maintaining a manageable number of independent variables in the HLM models, and (c) providing a powerful test of the effect of ESOL levels on achievement gains and gaps.

In HLM analyses described below, the independent variables were used at the level of the classroom. All independent variables that were collected at the level of the student (i.e., gender, ethnicity, ESE, ESOL, retention) were converted to proportions that were reflective of the classroom-level values of the variables. For example, the independent variable of gender was converted to the proportion of students in the class who were coded as being male. A list of the classroom-level independent variables used in HLM analyses is provided in Table 2.

A series of analyses were conducted to examine the gains in science achievement from pretest to posttest (Research Question 1) and the extent to which the gains depended on ESOL level and retention as measures of English proficiency and literacy (Research Question 2). First, descriptive statistics were obtained for the pretest, posttest, and gain scores across all students of the treatment group, and as a function of each of the independent variables described above. Second, to test whether gains existed and whether gaps in achievement changed from pretest to posttest as a function of English proficiency and literacy, an HLM analysis was conducted using an HLM statistical package (Raudenbush, Bryk, Cheong, & Congdon, 2000). In each model, we indicated the individual by i and the classroom by j , such that $GAIN_{ij}$ corresponds to the gain score for the i th individual in the j th classroom. The HLM analysis contained two primary models: a level-1 model and a level-2 model. The level-1 model expressed each individual's gain score as a function of the

Table 2

Summary of the classroom-level independent variables used in the HLM analyses

Independent Variable	Abbreviated Name	Description
Gender	GEN	Proportion of students in the class who were coded as being male
Ethnicity	ETH	Proportion of students in the class who were coded as being Black
ESE	ESE	Proportion of students in the class who were coded as ESE
ESOL	ESOL	Proportion of students in the class who were coded as ESOL levels 1, 2, 3, or 4
Retention	RET	Proportion of students in the class who were coded as being retained at least once
Treatment condition	TRT	1 = treatment group versus 0 = comparison group

ESE, exceptional student education; ESOL, English to speakers of other languages.

classroom mean gain score (b_{0j}) and error (r_{ij}). The level-2 model specified the classroom mean gain score (b_{0j}) as a function of the mean value of the individual classroom-level mean gain scores (b_{00}), a series of independent variables related to the demographic composition of the classroom (see Table 2), and error (u_j). The resulting level-1 and level-2 models can be displayed as:

$$\text{Level-1 model: } \text{GAIN}_{ij} = b_{0j} + r_{ij}$$

$$\text{Level-2 model: } b_{0j} = b_{00} + b_{1j}(\text{GEN}) + b_{2j}(\text{ETH}) + b_{3j}(\text{RET}) + b_{4j}(\text{ESE}) + b_{5j}(\text{ESOL}) + u_j$$

The level-2 model is often referred to as a mean-as-outcome model (Raudenbush & Bryk, 2002) because it is in essence a regression model using the classroom mean gain score as the outcome variable.

The only measures of achievement obtained for both the treatment and comparison groups were the strand scores on the statewide mathematics test (Research Question 3). As a result, analyses of these measures comprise the primary mechanism by which a direct comparison of the achievement of the treatment and comparison groups can be made. For each strand score, a series of analyses was conducted that parallel those for the science test scores described above. First, to obtain an overall picture of the strand scores, descriptive statistics were computed as a function of the treatment condition (treatment versus comparison) and each of the independent variables within the treatment and comparison groups. Second, to examine the extent to which between-classroom differences in mean strand score are attributable to the treatment condition and/or the demographic characteristics of the classroom, HLM analyses were conducted. Again, we indicated the individual by i and the classroom by j , such that Y_{ij} corresponds to the strand score for the i th individual in the j th classroom. The respective level-1 and level-2 models can be displayed as:

$$\text{Level-1 model: } Y_{ij} = b_{0j} + r_{ij}$$

$$\begin{aligned} \text{Level-2 model: } b_{0j} = & b_{00} + b_{1j}(\text{GEN}) + b_{2j}(\text{ETH}) + b_{3j}(\text{RET}) + b_{4j}(\text{ESE}) \\ & + b_{5j}(\text{ESOL}) + b_{6j}(\text{TRT}) + u_j \end{aligned}$$

Note that this level-2 model includes the variable coding for the treatment and comparison groups (TRT). Comparing the relative fit of the level-2 model with and without the treatment

variable addresses the extent to which the treatment variable explains between-classroom differences in mean scores on the particular strand of the statewide mathematics test. Because there are five strand scores under investigation, a separate set of HLM analyses is conducted for each strand.

Results

The primary purpose of this study involves science achievement of ELL students and students with limited literacy in English. To measure achievement gains (Research Question 1) and achievement gaps (Research Question 2) on the project-developed science test in the treatment group, we present the results by ESOL and retention (i.e., third-grade students who were retained due to failing scores on the statewide reading assessment from the previous year). The results by other demographic variables (i.e., gender, ethnicity, and ESE) are not reported. It is noted that these demographic variables were necessary for analysis purposes to reduce error variance and enhance statistical power. To measure students' performance on high-stakes testing, we compare the scores on the statewide mathematics test between the treatment and comparison groups (Research Question 3).

Science Achievement Gains and Gaps

Descriptive statistics for the mean pretest and posttest scores, as well as the gain scores, for all students and for each subgroup by ESOL and retention are presented in Table 3. For all students in the treatment group, the mean pretest score was 7.40 ($SD = 3.36$), the mean posttest score was 14.34 ($SD = 4.30$), and the mean gain score was 6.95 ($SD = 4.18$). The mean scores were lower for students classified as ESOL levels 1 to 4 than for students who had exited ESOL or never been in ESOL at both the pretest (6.55 versus 7.53) and posttest (12.39 versus 14.67). ESOL students at levels 1 to 4 had a lower mean gain score ($M = 5.84$) than students who had exited from ESOL or never been in ESOL ($M = 7.14$). Students who had been retained had lower mean scores than students who had not been retained at both the pretest (7.08 versus 7.48) and posttest (13.52 versus

Table 3
Descriptive statistics for science test scores (total possible score = 24 points)

Variable	Test	Subgroup	<i>N</i>	<i>M</i>	<i>SD</i>
All students	Pre		818	7.40	3.36
	Post		818	14.34	4.30
	Gain		818	6.95	4.18
ESOL	Pre	ESOL levels 1 to 4	118	6.55	3.40
		ESOL exited or non-ESOL	698	7.53	3.34
	Post	ESOL levels 1 to 4	118	12.39	4.62
		ESOL exited or non-ESOL	698	14.67	4.16
	Gain	ESOL levels 1 to 4	118	5.84	4.08
		ESOL exited or non-ESOL	698	7.14	4.17
RET	Pre	Not retained	632	7.48	3.30
		Retained	143	7.08	3.57
	Post	Not retained	632	14.58	4.31
		Retained	143	13.52	4.04
	Gain	Not retained	632	7.10	4.24
		Retained	143	6.44	3.84

Note: The ESOL information was missing for two students. The retention information was missing for 43 students.

14.58). The retained students had a lower mean gain score ($M = 6.44$) than students who had never been retained ($M = 7.10$).

The results of an HLM analysis modeling the classroom-level mean gain scores as a function of the five independent variables (see Table 2) are shown in Table 4. The table presents the obtained information for b_{00} , which represents the mean value of the classroom-level mean gain scores. The table also presents the results for the independent variables of ESOL and retention. The table provides the value of the associated coefficient (i.e., the slope), the standard error (*SE*) of the coefficient, the associated value of *t* statistic (the *t* statistic is obtained by dividing each coefficient by its standard error), degrees of freedom (*df*), and the *p*-value associated with the *t* statistic. To ensure adequate control of the Type I error rate, only coefficients for which $p < .01$ are interpreted as differing significantly from zero.

The overall mean gain score (denoted by b_{00} in the level-2 model described above) equaled 8.65 ($p = .000$), indicating that the overall gain (averaged across all classrooms) differed significantly from zero. Although the overall gain was significant, the magnitude of the gain was not dependent on ESOL or RET as evidenced by the nonsignificant coefficients for ESOL and RET. This latter finding indicates that the gains were not significantly different for the ESOL and retention groups, and thus the gaps in achievement by ESOL and retention did not display a statistically significant change from pretest to posttest.

Statewide Mathematics Test

The results for the measurement strand of the statewide mathematics test are presented in Table 5. The students who participated in the treatment had a higher mean score ($M = 5.00$) than the students in the comparison group ($M = 4.39$). In both the treatment and comparison groups, students who were classified as ESOL levels 1 to 4 had lower mean scores than those who had exited from ESOL or never been in ESOL, and students who had been retained had lower mean scores than those who had never been retained.

The results of an HLM analysis for the measurement strand are presented in Table 6. Of greatest interest to this study is the result obtained for the TRT independent variable (treatment versus comparison condition). The coefficient for TRT was 0.47, which was significantly different from zero ($p = .003$), indicating that TRT explained a significant proportion of the between-classroom differences in mean score on the measurement strand, after controlling for all other independent variables. The coefficient value of 0.47 for the TRT variable indicates that belonging to the treatment group is associated with an expected increase of 0.47 in the score of the measurement strand over that expected for the comparison group, after controlling for all other independent variables. In addition, the TRT variable explained 9% of the variance in classroom mean scores over and above that explained by the other independent variables (this value was obtained by examining the difference in the variance of the errors of prediction for the classroom mean score with and without the TRT variable included in the model).

Table 4
Results of HLM analysis for science test scores

Variable	Coefficient	<i>SE</i>	<i>t</i>	<i>df</i>	<i>p</i>
Mean gain score	8.65	1.06	8.16	35	.000
ESOL	-1.64	1.66	-0.99	35	.329
RET	-0.85	0.79	-1.07	35	.295

ESOL, English to speakers of other languages; RET, retention.

Table 5
Descriptive statistics for the measurement strand on the statewide mathematics test (total possible score = 8 points)

Variable	Treatment Condition	Subgroup	<i>N</i>	<i>M</i>	<i>SD</i>
All students	Treatment		942	5.00	1.91
	Comparison		966	4.39	2.02
ESOL	Treatment	ESOL levels 1 to 4	133	4.49	2.95
		ESOL exited or non-ESOL	789	5.87	2.96
	Comparison	ESOL levels 1 to 4	216	4.27	2.87
		ESOL exited or non-ESOL	724	5.18	2.99
RET	Treatment	Not retained	756	5.09	1.88
		Retained	186	4.65	1.96
	Comparison	Not retained	724	4.49	2.07
		Retained	242	4.07	1.81

Note: The ESOL information was missing for 20 treatment group students and 26 comparison group students. ESOL, English to speakers of other languages; RET, retention.

For the ESOL variable, the coefficient differed significantly from zero, indicating that students who had exited from ESOL or never been in ESOL performed significantly higher than students who were classified as ESOL levels 1 to 4 in both the treatment and comparison groups. For the retention variable, the coefficient did not differ significantly from zero, indicating that students who had been retained performed comparably to those who had never been retained in both the treatment and comparison groups.

Four additional HLM analyses were run, one for each of the remaining four strand scores (i.e., number sense, concepts, and operations; geometry and spatial sense; algebraic thinking; and data analysis and probability). In none of these cases did the TRT variable contribute significantly to the prediction of the respective classroom-level mean score.

Discussion and Implications

This study examined science achievement gains and achievement gaps at the completion of the first-year implementation of a teacher professional development intervention with ELL students in urban elementary schools. In this section, conclusions stemming from the results are discussed, followed by implications for our on-going intervention efforts and suggestions for further research.

Discussion

The main results of the study center around three findings. First, students in the treatment group displayed a statistically significant increase in science achievement. Second, there was no

Table 6
Results of HLM analysis for the measurement strand on the statewide mathematics test

Variable	Coefficient	<i>SE</i>	<i>t</i>	<i>df</i>	<i>p</i>
TRT	0.47	0.16	2.99	100	.003
ESOL	-1.00	0.35	-2.90	100	.004
RET	-0.35	0.23	-1.53	100	.125

RET, retention; ESOL, English to speakers of other languages; TRT, treatment condition.

statistically significant difference in achievement gains between students at ESOL levels 1 to 4 and students who had exited from ESOL or never been in ESOL. Similarly, there was no significant difference in achievement gains between students who had been retained and students who had never been retained. Third, the treatment group students showed a higher score on a statewide mathematics test, particularly on the measurement strand emphasized in the intervention, than the comparison group students.

The primary purpose of this study involves science achievement of ELL students and students with limited literacy in English (Research Questions 1 and 2). The results of the HLM analyses indicate that these students made significant achievement gains at the end of the school year. The results also indicate that students at ESOL levels 1 to 4 made achievement gains comparable to those who had exited from ESOL or never been in ESOL. Similarly, third-grade students, who had been retained due to failing scores on the statewide reading assessment from the previous year, made significant achievement gains comparable to those students who had never been retained. Achievement gains made by ESOL and retained students are noteworthy. The project-developed science test was administered in English to be consistent with the language of instruction. Although we encouraged teachers to make test accommodations, written assessments in English may underestimate the science knowledge of both ELL students and students with limited literacy in English. This, by extension, may underestimate the effectiveness of the intervention.

Another purpose of the study involves performance on high-stakes testing by ELL students and students with limited literacy in English (Research Question 3). Because statewide science assessment factoring into accountability did not exist, we used the statewide mathematics assessment, specifically the measurement strand that was emphasized in our intervention, as a proxy measure of high-stakes testing and accountability. The result of the HLM analysis indicates that the treatment group students performed significantly better than the comparison group students on the measurement strand. This result is noteworthy. The Measurement unit in our intervention was designed as basic concepts and skills for engaging in science inquiry; yet, the effectiveness was manifest on the measurement strand of statewide mathematics assessment.

Our professional development intervention is aimed at improving science and literacy achievement of ELL students in urban elementary schools. The literature indicates that an integrated approach to professional development that addresses ELL students' learning needs in English and content areas simultaneously would provide the greatest likelihood of success, because these multiple domains could mutually support one another (Lee & Luykx, 2005). Our intervention highlights English language and literacy and mathematics as part of science instruction. Despite its potential to promote student achievement as reported in this article, our intervention faces challenges because elementary teachers need extensive support to effectively teach science with ELL students in urban schools (for details, see Lee et al., 2006). Our intervention faces additional challenges as it involves scaling-up efforts with nonvolunteer teachers as a school-wide initiative.

Our professional development intervention involves more than curriculum units and teacher workshops. It also involves provision of supplies for science instruction in urban elementary schools that tend to have limited funding and resources (Spillane et al., 2001). Additionally, it involves ensuring instructional time for science in low-performing urban elementary schools where science tends to be ignored due to the urgency of developing basic literacy and numeracy (Lee & Luykx, 2005). Such factors should be considered for any professional development intervention to go to scale, especially in urban schools facing sanctions against poor academic performance on high-stakes testing and accountability (Settlage & Meadows, 2002). As science becomes part of high-stakes testing, it is essential that ELL students and other students who have

traditionally been underserved in the education system have opportunities to learn to think and reason scientifically while also performing well on high-stakes testing.

The results of the study provide support for the emerging literature on the positive impact of professional development interventions on science achievement of ELL students (Fradd et al., 2002; Hampton & Rodriguez, 2001; Lee et al., 2005). Not only did the students in the study develop an understanding of science concepts and inquiry, but they also performed well on high-stakes testing (Amaral et al., 2002). Discussion of ELL students' performance on high-stakes science testing is almost absent in the current literature, and the results of this study offer initial insights within the policy context increasingly driven by high-stakes testing and accountability across subject areas, including science that will become part of NCLB in 2007.

Implications

The results from the first-year implementation of our professional development point to areas for consideration in our ongoing intervention and research. As we continue our intervention, we will examine its impact on students' science and literacy achievement as measured by both project-developed tests and high-stakes tests. The results of our longitudinal research will contribute to the emerging knowledge base on science and English language and literacy with ELL students.

Different patterns in achievement gains among demographic groups raise questions for further consideration. Our intervention did not narrow achievement gaps for students at ESOL levels 1 to 4 and students who had been retained due to limited reading ability, although the gaps did not widen. The results indicate that more concerted efforts should be made to promote science learning while also supporting English language and literacy development of these students in our on-going intervention efforts to narrow achievement gaps.

In addition to examining the impact of the intervention on students' science achievement, the larger research also assessed literacy (writing) achievement using a writing prompt at the beginning and end of the school year. Students' writing samples are analyzed in terms of "form" (conventions, organization, and style/voice) and "content" (specific knowledge and understanding of science) in expository writing. Literacy achievement, coupled with science achievement, will provide insights to improve our on-going intervention efforts (see Amaral et al., 2002; Lee et al., 2005).

The results of science achievement need to be interpreted in relation to student experiences in science lessons, in that the professional development intervention's potential impact on student achievement was mediated by teachers' classroom practices with their students (briefly described above). The larger research examines teachers' knowledge and practices in integrating science with English language development for ELL students (Lee et al., 2006). Furthermore, the larger research examines the relationships between the fidelity of implementation in classroom practices and ELL students' achievement in science and literacy (writing).

In a longitudinal design, our professional development intervention is designed to improve elementary teachers' knowledge and practices in teaching science while promoting English language development of ELL students in urban schools. Our on-going intervention is assessed continuously based on multiple data sources, including student achievement outcomes. Teacher change, in turn, will have cumulative effects on student achievement over the years. In addition to examining the impact of the intervention on teacher change and student achievement, respectively, further research will examine the relationship between teacher change and student achievement as a result of the intervention (Fishman, Marx, Best, & Tal, 2004; Supovitz, 2001).

A primary motivation for our research is to improve science achievement of ELL students, especially in the context of the impending high-stakes testing and accountability policy in science in the state and the nation. When the third-grade students in this study advance into fifth grade in 2006, they will be the first cohort of students for whom the statewide science assessment will factor into school accountability. Over the years, our research will address two overarching questions: (a) how to promote ELL students to learn science while also developing English proficiency, and (b) how to promote ELL students to think and reason scientifically while also performing well on high-stakes science testing. The results, based on school-wide implementation of the professional development intervention with grades 3 through 5 teachers and their students in urban elementary schools, will provide insights about promoting science and literacy achievement of ELL students. Additionally, our on-going intervention and research will lead to better understanding about curriculum development, teacher professional development, and classroom practices leading to science and literacy achievement of all students.

Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the position, policy, or endorsement of the funding agency. The authors acknowledge the valuable feedback from Cory Buxton and Jane Sinagub.

References

- Amaral, O.M., Garrison, L., & Klentschy, M. (2002). Helping English learners increase achievement through inquiry-based science instruction. *Bilingual Research Journal*, 26, 213–239.
- Casteel, C.P., & Isom, B.A. (1994). Reciprocal processes in science and literacy learning. *The Reading Teacher*, 47, 538–545.
- Cuevas, P., Lee, O., Hart, J., & Deaktor, R. (2005). Improving science inquiry with elementary students of diverse backgrounds. *Journal of Research in Science Teaching*, 42, 337–357.
- Fishman, B., Marx, R., Best, S., & Tal, R. (2004). Linking teacher and student learning to improve professional development in systemic reform. *Teaching and Teacher Education*, 13, 43–76.
- Fradd, S.H., Lee, O., Sutman, F.X., & Saxton, M.K. (2002). Materials development promoting science inquiry with English language learners: A case study. *Bilingual Research Journal*, 25, 479–501.
- Hampton, E., & Rodriguez, R. (2001). Inquiry science in bilingual classrooms. *Bilingual Research Journal*, 25, 461–478.
- Kennedy, M. (1998). Form and substance in inservice teacher education (Research Monograph No. 13). Madison, WI: University of Wisconsin, National Institute for Science Education.
- Knapp, M.S., & Plecki, M.L. (2001). Investing in the renewal of urban science teaching. *Journal of Research in Science Teaching*, 38, 1089–1100.
- Lee, O. (2005). Science education and English language learners: Synthesis and research agenda. *Review of Educational Research*, 75, 491–530.
- Lee, O., Deaktor, R.A., Hart, J.E., Cuevas, P., & Enders, C. (2005). An instructional intervention's impact on the science and literacy achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching*, 42, 857–887.
- Lee, O., & Fradd, S.H. (1998). Science for all, including students from non-English language backgrounds. *Educational Researcher*, 27, 12–21.

Lee, O., LeRoy, K., Thornton, C., Adamson, K., Maerten-Rivera, J., & Lewis, S. (in press). Teachers' perspectives on a professional development intervention to improve science instruction among English language learners. *Journal of Science Teacher Education*.

Lee, O., Lewis, S., Adamson, K., Maerten-Rivera, J., & Secada, W.G. (2006). Urban elementary school teachers' knowledge and practices in teaching science to English language learners. Manuscript submitted for publication.

Lee, O., & Luykx, A. (2005). Dilemmas in scaling up innovations in science instruction with nonmainstream elementary students. *American Educational Research Journal*, 42, 411–438.

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

National Center for Education Statistics. (1999). *Teacher quality: A report on the preparation and qualifications of public school teachers*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning*. Washington, DC: National Academy Press.

No Child Left Behind Act of 2001. Public Law No. 107–110, 115 Stat. 1425. (2002).

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

Raudenbush, S.W., Bryk, A.S., Cheong, Y., & Congdon, R.T. (2000). *HLM 5: Hierarchical linear and nonlinear modeling*. Chicago, IL: Scientific Software International.

Settlage, J., & Meadows, L. (2002). Standards-based reform and its unintended consequences: Implications for science education within America's urban schools. *Journal of Research in Science Teaching*, 39, 114–127.

Spillane, J.P., Diamond, J.B., Walker, L.J., Halverson, R., & Jita, L. (2001). Urban school leadership for elementary science instruction: Identifying and activating resources in an undervalued school subject. *Journal of Research in Science Teaching*, 38, 918–940.

Stoddart, T., Pinal, A., Latzke, M., & Canaday, D. (2002). Integrating inquiry science and language development for English language learners. *Journal of Research in Science Teaching*, 39, 664–687.

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

Teachers of English to Speakers of Other Languages. (1997). *ESL standards for pre-K-12 students*. Alexandria, VA: Author.

Wong-Fillmore, L., & Snow, C. (2002). *What teachers need to know about language*. Washington DC: Center for Applied Linguistics.