EFFECT OF A MULTIYEAR INTERVENTION ON SCIENCE ACHIEVEMENT OF ALL STUDENTS INCLUDING ENGLISH LANGUAGE LEARNERS

ABSTRACT

This study was part of the Promoting Science among English Language Learners (P-SELL) efficacy study, a research and development project that implemented a curricular and professional development intervention to improve science achievement of English Language Learners (ELLs) in urban elementary schools. The study used a cluster randomized control trial (RCT) design with 31 treatment schools and 32 control schools over a 3-year period. This study used a series of hierarchical generalized linear models (HGLMs) to examine science proficiency as measured by state high-stakes science assessment. The results indicated differences in science proficiency between the treatment and control groups in Years 2 and 3, but not in Year 1. The results demonstrate that an educational intervention based on the synergy of science and English language and literacy for ELLs is effective at increasing students' science achievement. However, there may be a lag in improved student achievement as teachers gain familiarity with the new curriculum and adjust their teaching practices over time.

Jaime Maerten-Rivera
STATE UNIVERSITY OF NEW YORK AT BUFFALO

Soyeon Ahn
UNIVERSITY OF MIAMI

Kimberly Lanier
MIAMI DADE COLLEGE

Jennifer Diaz
THE EVERGLADES FOUNDATION, INC.

Okhee Lee
NEW YORK UNIVERSITY

 0013-5984/2016/11604-0004$10.00

GIVEN the increasing urgency in the call for high academic standards in science for all students, science education reform efforts have focused explicitly on nonmainstream students, including English Language Learners (ELLs; NRC, 1996, 2000). Science education has become more prominent since the 2007–2008 school year when it was included in accountability policies as part of the No Child Left Behind Act of 2001. Although science is not part of adequate yearly progress (AYP) calculations, some states include science in their state accountability systems (comprehensive and systematic information across the states is not currently available). The Next Generation Science Standards will further push science education reform. At the same time, standards for teaching ELLs specifically target academic language proficiency in four core content areas, including science, as a central goal (TESOL, 2006). Thus, educational interventions that promote academic learning for all students, including ELLs, are greatly needed.

Over the course of its 3-year period from 2010–2011 through 2012–2013, the Promoting Science among English Language Learners (P-SELL) project, of which this study was a part, developed and implemented a curricular and professional development intervention to improve science achievement of all students, particularly ELLs, in urban elementary schools. The study focused on science proficiency as measured by state high-stakes science assessment in fifth grade, which was the grade at which science counted toward school accountability. The study examined the following research questions in order to examine the effect of the intervention:

1. Was there a difference in science proficiency between the treatment and control groups each year (i.e., Year 1, Year 2, and Year 3) after controlling for background characteristics (i.e., gender, ethnicity, socioeconomic status [SES], exceptional student education [ESE], and English for Speakers of Other Languages [ESOL])?
2. Was the effect of the intervention immediate (i.e., Year 1), or was the effect of the intervention delayed until Year 2 or Year 3?

This study contributes to the emerging literature on educational interventions to promote science learning of all students, especially ELLs (Buxton & Lee, 2014; Janzen, 2008; Lee, 2005). The study designed, implemented, and examined a large-scale curricular and professional development intervention that capitalized on the intersections between science and English language and literacy for all students and ELLs in particular. Over the 3-year duration of the study, the intervention was implemented with three cohorts (i.e., multiyear replication) of fifth-grade students to look for trends in science proficiency.

Conceptual Framework

The growing population of ELLs presents challenges to teachers who must enable these students to acquire academic knowledge across content areas while simultaneously developing the students' English language and literacy skills. There is now broad agreement in terms of effective teaching practices to support robust science and language learning, and efforts are emerging to implement these practices on a large scale and evaluate their effectiveness (Buxton & Lee, 2014).

Integration of Science Instruction and Language Instruction with ELLs

In much of the early literature on effective science instruction with ELLs, researchers highlighted the importance of engaging students in hands-on activities to make science concrete and experiential. More recently, researchers highlight the need to integrate science inquiry with an explicit focus on language and literacy development in English. Researchers also highlight the use of ELLs' home language and culture as an instructional support to enhance science learning.

Reform-oriented practices in science. Science education reform highlights that effective science instruction promotes students' scientific understanding and inquiry. Students should develop a deep and complex understanding of science concepts, make connections among concepts, and apply concepts in explaining natural phenomena or real-world situations (NRC, 1996, 2007, 2011). Also, students should engage in science inquiry related to the practice of science (NRC, 2000).

To promote students' science inquiry and understanding, teachers need to know subject matter content and content-specific teaching strategies (Abell, 2007; Kennedy, 1998; van Driel, Berry, & Meirink, 2014). They should recognize the role of prior knowledge, particularly students' misconceptions, in shaping students' understanding. In addition to engaging students in science inquiry and fostering initiative in inquiry (NRC, 2000), teachers need to enable students to recognize problematic and incomplete information, make reasoned and well-supported arguments, justify claims based on evidence, share and negotiate ideas, and construct collective meanings about science (NRC, 2007). Teachers with more science knowledge produced higher science achievement from their students (Diamond, Maerten-Rivera, Rohrer, & Lee, 2014; Heller, Daeler, Wong, Shinohara, & Miratrix, 2012). In addition, inquiry-based science instruction resulted in higher science achievement than commonplace science instruction (Blanchard, Southerland, & Granger, 2009; Wilson, Taylor, Kowalski, & Carlson, 2010).

English language and literacy in science with ELLs. While the value of hands-on inquiry approaches continues to be an important component of science instruction for ELLs, a current emphasis has been on creating integrated models of instruction that promote science inquiry while simultaneously focusing on language development in English (Buxton & Lee, 2014; Janzen, 2008; Lee, 2005; Rosebery & Warren, 2008). First, effective teachers highlight various strategies for developing content area literacy (reading and writing), including activation of prior knowledge, comprehension of expository science texts, awareness of academic language functions in relation to science process skills, practice with scientific genres of writing, and use of graphic organizers. Second, effective teachers utilize second-language pedagogies (ESOL strategies) and strategies typical of contextualized experiential approaches, including hands-on activities, realia, purposeful activities, and multiple examples of language in various contexts. Third, effective teachers facilitate ELLs' participation in classroom discourse to enhance students' understanding of academic content. They are sensitive to the various levels of students' developing language proficiency, adjust the level and mode of communication, and use multiple modes of representation (gestural, oral, pictorial, graphic, and textual).

Fourth, effective teachers focus on students' home language as an instructional support (Goldenberg, 2013). They use science terms in students' home language, highlight cognates between English and the home language, allow code-switching, and encourage bilingual students to assist less English proficient students in their home language. Finally, effective teachers elicit students' "funds of knowledge" from home and community contexts related to science topics (González, Moll, & Amanti, 2005) and use students' cultural artifacts and community resources in ways that are both academically meaningful and culturally relevant (Solano-Flores & Nelson-Barber, 2001).

Curricular and Professional Development in Science with ELLs

The emerging literature indicates key components of effective professional development in general. First, effective professional development considers both core and structural features (Desimone, 2009; Garet, Porter, Desimone, Birman, & Yoon, 2001; Penuel, Fishman, Yamaguchi, & Gallagher, 2007). Core features of effective professional development include (*a*) focus on content knowledge and how students learn that content, (*b*) opportunities for teachers to engage in active learning, and (*c*) coherence with other activities for teacher learning and development. In addition, structural features include (*d*) sufficient duration in terms of both the number of contact hours and the span across the calendar year, and (*e*) collective participation of teachers from the same school, department, or grade level.

Second, effective professional development enables teachers to fully realize the intentions of the curriculum and utilize the curriculum as scaffolds to promote teacher learning in addition to student learning (Remillard, 2005). Davis and Krajcik (2005) proposed the notion of "educative curriculum materials" to promote changes in teachers' science knowledge and practices in specific instances of instructional decision making and also to help teachers develop more general knowledge that they can apply in new situations (Davis et al., 2014; Drake, Land, & Tyminski, 2014).

Finally, effective professional development for teachers of ELLs enables them to explicitly connect science learning with literacy development for ELLs, such that rigorous content learning in science and other subjects provides an authentic context for gaining academic language skills (Janzen, 2008; Lee, 2005). In addition, teachers learn to capitalize on ELLs' linguistic experiences (e.g., home language) and cultural experiences (e.g., funds of knowledge) that can be used as intellectual resources (Warren, Ballenger, Ogonowski, Rosebery, & Hudicourt-Barnes, 2001).

Intervention Effect on Science Achievement of All Students, with a Focus on ELLs

Research on professional development interventions to promote science achievement of ELLs has begun to emerge in recent years (see the literature review by Buxton & Lee, 2014). In the context of inquiry-based instruction, studies have examined the impact of interventions on ELLs' science achievement. Overall, the results have shown promise for improving achievement gains as well as narrowing achievement gaps for ELLs.

August, Branum-Martin, Cardenas-Hagan, and Francis (2009) assessed the effectiveness of a professional development intervention consisting of curriculum materials, teacher workshops, and weekly mentoring. The study involved 10 sixth-grade science teachers in five middle schools in a large school district. For each teacher, two class sections were randomly assigned to the treatment group that focused on inquiry-based science, and two class sections were randomly assigned to the control group that used the district curriculum. The students in the treatment group demonstrated significantly higher achievement gains than those in the control group for both science knowledge and vocabulary.

Lee, Deaktor, Enders, and Lambert (2008) implemented a professional development intervention aimed at promoting science and writing achievement for culturally and linguistically diverse elementary students. Through the provision of curricular materials and teacher workshops, the intervention emphasized the integration of (*a*) inquiry-based science, (*b*) English language and literacy, and (*c*) students' home language and culture. Students demonstrated significant achievement gains, and achievement gaps by demographic subgroups narrowed or remained consistent.

Recently, professional development researchers have noted the difficulty of evaluating interventions because effects may be delayed due to the fact that time is needed to implement professional development activities. When teachers engage in professional development, student outcomes may not be evident for 2 or 3 years (Kreider & Bouffard, 2006). The claim of delayed effects of professional development was substantiated in a study examining passing rates on high-stakes science assessments following teacher professional development; differences in passing rates between students of participating teachers and nonparticipating teachers were not significant during the first 2 years but were significant during the third and fourth years (Silverstein, Dubner, Jon, Glied, & Loike, 2009). Wayne, Yoon, Zhu, Cronen, and Garet (2008) discuss the need for randomized control experiments to evaluate professional development and its effect on student outcomes, including examination of how long it takes for improved instruction to lead to detectable increases in student outcomes (i.e., whether professional development affect student outcomes during the year in which the professional development is received or whether there is a delayed impact).

The literature on teacher professional development interventions highlights that science inquiry consistently promotes ELLs' science achievement and sometimes reduces achievement gaps between ELLs and non-ELLs. Building on the emerging literature, our curricular and professional development intervention (described below) identified key features of science and English language and literacy and then capitalized on their intersections. The study examined the intervention's effect on science proficiency of elementary students involving three cohorts over 3 years.

The P-SELL Intervention

In our intervention, curriculum and professional development were designed to complement each other with the goal of improving teachers' science knowledge and teaching practices according to state science standards for all students, and ELLs in particular, in a high-stakes assessment and accountability policy context. The intervention was grounded in three areas: (*a*) reform-oriented practices to promote students' science inquiry and understanding, (*b*) state science standards and high-stakes science assessment, and (*c*) science instruction for diverse student groups with a focus on ELLs. Throughout the 3 years of the intervention, the control schools implemented the district-adopted science curriculum.

Curriculum in the Intervention

A comprehensive stand-alone year-long fifth-grade science curriculum encompassing the nature of science, earth and space science, life science, and physical science was developed for this intervention. The intervention provided each treatment teacher with a teachers' guide, a class set of consumable student books, and supplies. The class set of student books was provided each school year, and consumable supplies were replenished at the start of each subsequent school year. Supplementary materials were made available via CD and also accessible online at the project's website.

The curriculum highlighted a standards-based and inquiry-oriented approach for all students, especially ELLs, with a focus on three key features. First, the curriculum used a standards-based approach by aligning with state science standards and high-stakes science assessment administered at fifth grade. The state science standards consisted of 18 "big ideas" according to four "bodies of knowledge" including the nature of science, earth and space science, life science, and physical science. The curriculum was organized around these big ideas. Each chapter started with identification of the science standards and benchmarks addressed. Furthermore, each hands-on inquiry activity, reading passage, and writing section designated the science standard(s) and benchmark(s) addressed. Special consideration was given to science assessment item specifications in order to ensure alignment between the curriculum and the benchmarks of the state science assessment.

Second, the curriculum highlighted an inquiry-oriented approach. Science inquiry was emphasized both as a goal of science learning and as a means through which students developed scientific understanding of the big ideas in the state science standards. The student book was designed to move progressively from teacher-directed to student-directed inquiry. By providing more structure in earlier chapters, and a more open-ended approach in later chapters, the curriculum encouraged student initiative and exploration. The student book also emphasized key science concepts and big ideas. In addition, the student book included activities for student discussion and opportunities for sharing their reasoning.

Finally, the curriculum addressed the needs of ELLs by providing explicit guidance to promote understanding of science concepts and inquiry and to support English language development. Each chapter in the student book started with key science terms in English, Spanish, and Haitian Creole, the three primary languages spoken by students of the participating school district. Each chapter concluded with an expository text summarizing key science concepts in English, and translations into Spanish, Haitian Creole, and Russian were available on the project website. The student book used multiple modes of representation in textual and graphic formats (e.g., students wrote their ideas in the student book) and oral and aural forms (e.g., students discussed their ideas in small and whole groups).

The teachers' guide was designed to assist teachers with curriculum implementation. The front matter of the teachers' guide offered explanations on how the curriculum was designed to enable students to master the state science standards, why science inquiry was key to enabling students to understand the big ideas in the state science standards, how teachers could guide students toward student-initiated inquiry, and how teachers could support language and literacy development of all students, especially ELLs. For each chapter, following the science inquiry activities, the teachers' guide provided science background information and explanations for the questions under investigation and related natural phenomena, with an emphasis on students' common misconceptions and learning difficulties. In addition, the teachers' guide provided content-specific teaching strategies for each chapter. For example, it offered suggestions about how to set up and implement hands-on activities, along with cautions about potential problems and how to respond to such situations. It offered suggestions for different levels of guidance and scaffolding by using additional activities for students who needed support for content mastery as well as enrichment activities for students who needed challenge beyond content mastery. It also included language development strategies and an extensive list of web-based resources.

At the conclusion of Year 1, the student book and teachers' guide were revised to reflect new science standards adopted by the state. While the focus on key features of the intervention (i.e., standards-based, inquiry-oriented, and ELL-focused) remained intact, the revision addressed the changes in science content in the new standards and the order in which the science content was organized according to the new standards. In addition, some hands-on activities were modified in response to teacher feedback to increase ease of implementation.

Teacher Workshops in the Intervention

Teacher workshops focused on curriculum implementation by enabling teachers to realize the intentions of "educative curriculum materials" (Davis & Krajcik, 2005; Davis et al., 2014; Drake et al., 2014). The workshops incorporated critical features of effective professional development (Desimone, 2009; Garet et al., 2001; Penuel et al., 2007). The workshops focused on teachers' knowledge of science and reform-oriented teaching practices to promote students' science inquiry and understanding (i.e., content focus). Teachers were actively engaged in hands-on science inquiry and planning for classroom implementation (i.e., active learning). The workshops highlighted how the curriculum was aligned with state science standards and high-stakes science assessment, while also supporting literacy and English-language development (i.e., coherence). The workshops were offered during the summer and throughout the school year over the 3-year period (i.e., duration). All fifth-grade science teachers from the treatment schools participated (i.e., collective participation).

During Year 1, treatment teachers received 5 days of workshops, including 3 days in the summer shortly before school started, 1 day in January, and 1 day in May. Teachers were given a stipend for attending the summer workshops, and schools received substitute payments during school days. The workshops focused on familiarizing teachers with the curriculum by demonstrating strategies for hands-on inquiry activities, as well as highlighting the curriculum's alignment to the state science standards and high-stakes science assessment. The intervention helped teachers recognize how students' science inquiry abilities were related to the state science standards and, thus, could enhance performance on the high-stakes science assessment. Additionally, language development strategies were introduced, as they were embedded in science inquiry and understanding.

During Year 2, treatment teachers were offered 5 days of workshops that followed the same schedule as in Year 1. At the summer workshops, teachers new to the intervention were introduced to the curriculum just like the teachers during Year 1. With teachers returning to the intervention, changes to the new state science standards and the new high-stakes science assessment aligned with the new standards were highlighted. Building on professional development during the first year, the summer workshop with returning teachers highlighted further professional growth in science instruction (e.g., student initiative in science inquiry, common misconceptions), integration of science with English-language development, and incorporation of students' home language and culture in science instruction. At the January workshop, both new and returning teachers became familiar with the curriculum that was revised to align with the new state science standards and high-stakes assessment, while returning teachers shared their experiences from the previous year. The May workshop with both the new and returning teachers focused on unpacking science inquiry—highlighting opportunities for more open-ended questioning in science instruction, providing activities to help teachers become more comfortable and willing to implement inquiry, making teachers aware of what is a good student response, providing constructive feedback to students, and developing and using a rubric for assessing science inquiry.

During Year 3, returning teachers participated in a 1-day summer workshop. They were provided with students' work samples to discuss how to promote student-centered inquiry and how to pay explicit attention to language development strategies for ELLs. Then, they used a rubric to assess students' work samples with regard to science inquiry and language development. Teachers new to the intervention participated in a 3-day summer workshop to familiarize themselves with the curriculum for the school year. At the completion of the summer workshop, both the returning and new teachers engaged in collaborative planning with other teachers at their schools. At the meeting in May, both the new and returning teachers convened to offer feedback on the intervention.

Curriculum of the Control Group

Teachers in control schools implemented the district-adopted science curriculum. The fifth-grade textbook consisted of 18 chapters in four units: life science, physical science, earth science, and human body. The textbooks had the following features. First, the textbook provided four kits of science supplies (one kit for each unit) for each classroom for a total of 40 hands-on activities that included (*a*) 18 "directed inquiry" activities in a one-page format in the beginning of each chapter for demonstration purposes, (*b*) 18 "guided inquiry" activities in a two-page format at the end of each chapter that focused on following procedures, gathering data, and interpreting the data, and (*c*) four "full inquiry" activities on several pages at the end of each unit that involved a complete inquiry process. Second, the textbook linked science to other subjects, including literacy. It offered suggestions for reading and writing activities and used graphic devices including photos, diagrams, tables, and charts.

There were major differences between the district-adopted curriculum and the intervention curriculum. First, despite the large number of hands-on activities in the district-adopted curriculum, students engaged in a comprehensive inquiry process only during the four full inquiry activities. Also, hands-on activities were often disconnected from the science concepts of the chapter. Additionally, the district-adopted curriculum followed the traditional approach to presentation of science information. In contrast, the intervention curriculum highlighted hands-on activities as a means to engage students in science inquiry and to foster scientific understanding. Second, in the district-adopted curriculum, literacy activities were often disconnected from the science concepts of the chapter. It was text-heavy and required fairly high reading comprehension abilities to process the information. There was no clear evidence of strategies for ELLs. In contrast, the intervention curriculum integrated literacy as an essential component of science instruction, with special consideration of language development strategies for ELLs.

Method

Research Setting

This study took place in a large urban school district in the southeastern United States with diverse student and teacher populations. During the first year (2010–2011) of the project, the ethnic makeup of the student population in the school district was approximately 65% Hispanic, 24% Black non-Hispanic (including Haitian and Caribbean Islanders), 9% White non-Hispanic, and 2% Other. Across the school district, 72% of elementary students received free or reduced-price lunch, 19% participated in ESOL programs, and 11% participated in ESE programs excluding gifted education programs.

Research Design

At the time when schools were randomly selected to participate, there were 238 elementary schools in the district. Initially 23 schools were removed from the pool due to participation in alternate district interventions, and nine schools were removed because they had participated in a previous version of our study. This resulted in a final pool of 206 eligible schools. From this pool, 64 schools were randomly selected to participate in the study. The 64 schools were then randomly assigned with 32 to the treatment group and 32 to the control group.

In the year prior to the intervention, one treatment school implemented a science curriculum developed by the teachers at the school. During Year 1, the teachers were implementing some lessons from the intervention and other lessons from the teacher-developed curriculum. Since there was concern regarding fidelity to the intervention curriculum, the district, principal, and teachers decided that the school would not participate in the intervention after Year 1. This school was typically high performing, and the data from Year 1 showed that the school performed differently from other schools across science topics (e.g., higher scores in most areas). In order to accurately assess our intervention, and not another specially designed curriculum, the school was dropped from all analyses.

Therefore, a total of 31 treatment schools and 32 control schools participated over the 3 years. In the 2009–2010 school year, of the 31 treatment schools, 22 received letter grades of an "A," 5 a "B," 4 a "C," and 0 a "D." Of the 32 control schools, 21 received letter grades of an "A," 5 a "B," 5 a "C," and 1 a "D."

Teacher and Student Participants

There were a total of 351 fifth-grade science teachers who participated in the study over the 3 years, including 176 in the treatment group and 175 in the control group. Table 1 displays the number of teachers in both the treatment and control groups for each of the 3 years along with the number and percent of teachers who participated for their first, second, or third year in the intervention. Throughout the article, we use "T" to identify the treatment group and "C" to identify the control group.

Prior to beginning the intervention, teachers were asked to complete a brief survey about their background, including gender, ethnicity, native language(s) spoken, highest degree obtained, number of college science courses taken, number of years teaching, and ESOL endorsement. Table 2 presents the teachers' demographic and professional information. The two groups of teachers were comparable. The majority were female (T = 82%, C = 84%) and Hispanic (T = 55%, C = 54%). Most teachers spoke English as a native language (T = 83%, C = 79%), yet many identified Spanish as their native language (T = 44%, C = 53%). The highest degree obtained for most teachers was a bachelor's degree (T = 49%, C = 47%), while many had a master's degree (T = 44%, C = 42%). Most were endorsed to teach ESOL students (T = 92%, C = 90%). It should be noted that the ethnic backgrounds of the teacher sample reflected both the teacher population and the student population in the school district.

The school district provided a roster of the students taught by each science teacher at each school. The district also provided a database with the students' demographics and state assessment data. The students who were not included in the roster and state assessment data left the intervention school prior to the end of

Table 1. Teacher Participation over 3 Years

	Year 1 *n* (%)		Year 2 *n* (%)		Year 3 *n* (%)	
Participation	T	C	T	C	T	C
1 year	118 (100)	110 (100)	36 (33)	37 (37)	22 (22)	28 (28)
2 years	–	–	73 (67)	64 (63)	23 (23)	23 (23)
3 years	–	–	–	–	53 (54)	48 (49)
Total	118	110	109	101	98	99

Table 2. Teacher Demographic and Professional Background (n = 351)

Sample Characteristic	T (n = 176) (%)	C (n = 175) (%)
Gender:		
Male	18	13
Female	82	84
Missing	0	2
Ethnicity:		
Hispanic	55	54
Black non-Hispanic	23	19
White non-Hispanic	19	19
Haitian	2	1
Other	1	4
Missing	0	3
Native language(s):[a]		
English	83	79
Spanish	44	53
Haitian Creole	2	4
French	1	3
Other	1	3
Highest degree:		
Bachelor's	49	47
Master's	44	42
Specialist	4	6
Doctorate	1	1
Missing	2	5
ESOL endorsed:		
Yes	92	90
No	8	10

[a] Multiple native languages could be selected.

the school year or entered later in the school year, and thus represent attrition to the intervention. Table 3 presents the number of students included in the analyses for each year, along with the number of students lost to attrition for both the treatment and control groups.

Some schools were departmentalized where a science teacher taught multiple classes of fifth-grade science. In Year 1, the average number of students per teacher was 31.39 (SD = 17.81) in the treatment group and 31.85 (SD = 17.96) in the control group. In Year 2, the average number of students per teacher was 30.92 (SD = 16.66) in the treatment group and 32.68 (SD = 17.87) in the control group. In Year 3, the average number of students per teacher was 31.82 (SD = 19.56) in the treatment group and 33.66 (SD = 17.68) in the control group.

Table 3. Student Sample Each Year

	Year 1 n (%)		Year 2 n (%)		Year 3 n (%)	
	T	C	T	C	T	C
Analysis sample	3,202 (94)	3,581 (97)	3,093 (93)	3,376 (95)	2,958 (92)	3,333 (96)
Attrition	206 (6)	116 (3)	233 (7)	169 (5)	265 (8)	147 (4)
Total	3,408	3,697	3,326	3,545	3,223	3,480

Table 4 presents the demographics of the fifth-grade students from the treatment and control groups. For both groups across the 3 years, the majority of students were classified as Hispanic (about 65%), followed by Black (about 25%). A very high percent was classified as receiving free or reduced-price lunch (between 77% and 82%); each year the percent of students receiving free or reduced-price lunch was slightly higher in the control group than the treatment group. Nearly 50% of students in each group had exited from ESOL over 2 years prior or were never in ESOL, while the percent of students who had exited from ESOL within 2 years ranged from 31% to 42%, and the percent of students classified as ESOL levels 1–4 ranged from 12% to 23%. It should be noted that over the 3-year period there was an increase in the number of students classified as ESOL levels 1–4.

State High-Stakes Assessment and School Accountability

The state's school accountability system originated in 1999 and has been revised periodically. All public schools were assigned a "school grade" (A, B, C, D, or F) based on state high-stakes assessments. At the elementary level, students were assessed on reading and mathematics in grades 3–5, writing in grade 4, and science in grade 5. The percent of students scoring proficient on the fifth-grade science assessment counted for 1/8 of the school grade.

Beginning in the 2011–2012 school year (Year 2 of the study), ELLs with 1 full year of instruction in the United States at the time of testing were included in school performance calculations. In years prior, the date of entry to ESOL was used as the means of inclusion of ELLs in school performance calculations. This change

Table 4. Student Demographics by Group for 3 Years (Reported as Percentages)

	Year 1		Year 2		Year 3	
Variable	T (n = 3,202)	C (n = 3,581)	T (n = 3,093)	C (n = 3,334)	T (n = 2,958)	C (n = 3,333)
Gender:						
Male	50	51	52	52	50	51
Female	50	49	48	48	50	49
Ethnicity:						
Hispanic	66	65	65	65	66	66
Black	24	26	25	25	24	25
White non-Hispanic	8	7	8	8	9	8
Other	2	2	2	2	1	1
FRL:						
Received free or reduced-price lunch	77	80	77	81	77	82
ESE:						
Exceptional students	11	11	11	10	9	9
ESOL:						
ESOL levels 1–4	12	14	15	17	19	23
Exited ESOL within 2 years	39	42	36	38	31	31
Exited ESOL over 2 years ago or never in ESOL	49	44	49	45	50	46

led to ELLs being assessed, and their results counting toward school grades, at a lower level of proficiency. Students with disabilities (i.e., ESE) were required to take state assessments, and their results were included in the school grade calculations for proficiency in reading, math, writing, and science.

Science proficiency on the fifth-grade state assessment was the outcome variable, as other interventions have evaluated outcomes based on the percent of students deemed "passing" state tests (Silverstein et al., 2009; Weaver & Dick, 2009). Science proficiency, instead of science scale scores, was used to examine the intervention effect for several reasons. First, because accountability policies were based on percent proficient as opposed to scale scores, district and school administrators were interested in examining intervention effects in terms of percent proficient. Second, examination of percent proficient answered a different research question as compared to examination of scale scores and provided different results. In our intervention, while there were differences in scale scores between the treatment and control groups, differences in proficiency were more pronounced. Third, accountability policies are concerned with proficiency to aid low-performing students to proficiency as opposed to increasing the scores of average or above-average students, thus widening achievement gaps. Because a main goal of our intervention was to increase the performance of low-performing students, including ELLs, proficiency was a more suitable outcome. Finally, the science scale scores in the state changed in Year 3 of the study (see below for a description). Thus, the scale scores across the 3 years meant something different, whereas percent proficient provided the same information across the 3 years.

Because state science assessment in elementary schools was administered in fifth grade only, there was no science score from previous years to include as a covariate. To control for some of the variations in science proficiency, reading scale scores were used as a covariate. Research indicates that reading achievement is related to science achievement, possibly even more so than math achievement (Maerten-Rivera, Myers, Lee, & Penfield, 2010).

State science assessment. In Year 1 of the study, the state science assessment covered the topics of physical and chemical science, earth and space science, life and environmental science, and scientific thinking. The assessment included 45–55 multiple-choice questions. In 2006, the state reported that for all students across the state, the Cronbach's alpha was .87 for the science assessment. In Year 2 of the study, the state science assessment was changed to cover the areas of nature of science, earth and space science, physical science, and life science. The assessment included 60–66 multiple-choice questions.

In Years 1 and 2, the science scale ranged from 100 to 500. Based on scale scores, students were assigned achievement levels ranging from 1 to 5. Students with scores of 100–272 were level 1, 273–322 level 2, 323–376 level 3, 377–416 level 4, and 417–500 level 5. For the purpose of accountability, students classified as level 3, 4, or 5 were considered proficient. In Year 3 of the study, the science scale was changed to range from 140 to 260. Students with scores of 140–184 were level 1, 185–199 level 2, 200–214 level 3, 215–224 level 4, and 225–260 level 5. Again, students classified as level 3, 4, or 5 were considered proficient.

State reading assessment. In Year 1 of the study, the state reading assessment covered the topics of words and phrases in context; main idea, plot, and purpose;

comparisons and cause/effect; and reference and research. The reading assessment included 50–55 multiple-choice questions. In 2006, the state reported that for all students across the state, the Cronbach's alpha was .87 for the reading assessment. In Year 2 of the study, the state reading assessment was changed to cover the areas of vocabulary, reading application, literary analyses and informational text, and research process.

In Year 1, the reading scale ranged from 100 to 500. In Years 2 and 3, the reading scale was changed to range from 140 to 302 (note that this was 1 year prior to the change in science). In Year 1, the mean reading score for the treatment group was 306.36 ($SD = 58.71$), and the control group was 300.49 ($SD = 58.46$). In Year 2, the treatment group mean was 221.35 ($SD = 21.81$), and the control group mean was 220.17 ($SD = 21.82$). In Year 3, the treatment group mean was 219.37 ($SD = 20.77$), and the control group mean was 218.21 ($SD = 20.99$). It is noted that starting in Year 2, the mean was lower since the scale changed.

Data Analysis

The state determined the science proficiency levels of 1–5 for each student. For the purpose of the analyses in this article, the science scores were recoded into a dichotomous variable (SCIPROF), with levels 1 and 2 coded as 0, and levels 3, 4, and 5 coded as 1. Thus, a student received a 0 if s/he was classified as not proficient (i.e., did not pass) and a 1 if s/he was classified as proficient (i.e., did pass). To evaluate whether the odds of being proficient on state science assessment differed between students in the treatment group and those in the control group, a series of HGLMs (Raudenbush & Bryk, 2002), also referred to as hierarchical logistic models, were examined for each year of the intervention.

Analytic consideration. The state science assessment data consisted of students nested within teachers nested within schools. Although some teachers taught multiple classes (e.g., when the school was departmentalized), it was impractical to include the classroom nesting in the model because of the small number of science classes per teacher, with many teachers teaching only one science class in self-contained classrooms. When teachers taught multiple science classes, they typically taught in the same fashion across classes, and the demographics across classes tended to be similar; thus, we grouped the teachers' students together.

We first examined a fully unconditional three-level model for each year by nesting students within teachers within schools to determine the intraclass correlation coefficient (ICC; Snijders & Bosker, 1999), which estimated the variability attributed to the teacher level and the school level. The results for all 3 years were similar, with the school level accounting for less than 3% of the variance and the teacher level accounting for about 35% of the variance each year. Additionally, in each model, the estimates for the fixed effects at the school level were not statistically significant. As a result, we used a two-level model to examine the effect of the treatment on science proficiency with students nested within teachers.

Model process. To evaluate differences in the odds of students being proficient between the treatment and control groups, a series of two-level HGLMs were constructed. The level-1 model posited students' proficiency (not proficient = 0, proficient = 1) as a function of relevant demographic variables including (*a*) gender (GEN: female = 1, male = 0), (*b*) ethnicity (BLK: dichotomized as Black = 1 and non-Black = 0; WHT: dichotomized as White and Other = 1 and non-White, non-Other = 0), (*c*) free or reduced-price lunch status (FRL: 1 = yes, 0 = no), (*d*) ESE status (ESE: 1 = all ESE categories other than gifted, 0 = non-ESE or gifted), and (*e*) ESOL status (ESOL1–4: dichotomized as levels 1 to 4 = 1 and non-ESOL levels 1 to 4 = 0; ESOL5: dichotomized as ESOL level 5 = 1 and non-ESOL level 5 = 0). These level-1 predictor variables were included because previous research demonstrated their potential impact on science achievement (Lee, Maerten-Rivera, Penfield, LeRoy, & Secada, 2008; Maerten-Rivera et al., 2010), and because differences were noticeable on many of these variables in our sample (see Table 5). The dummy variables for ethnicity were created such that Hispanics, the group with the largest percentage in the sample, were used as the reference group to whom Blacks and Whites were directly compared. Similarly, the group of students who were never in ESOL or exited from ESOL over 2 years before was used as the ESOL reference group. Table 4 includes the descriptive information for student demographics. In addition, students' reading scale scores (READ) were included as a covariate, as described earlier. All level-1 predictors were entered into the model uncentered, with the exception of reading, which was group-mean centered.

The level-2 model posited teacher-level effects (e.g., teacher-level odds of being proficient) as a function of two teacher-level predictor variables: (*a*) mean reading assessment score across teacher (MREAD), which was grand-mean centered, and (*b*) the group to which the teacher was assigned (GROUP: 1 = treatment, 0 = control), which was entered into the level-2 model uncentered. The MREAD var-

Table 5. Students Classified as Proficient by Group for 3 Years (Reported as Percentages)

	Year 1		Year 2		Year 3	
Variable	T	C	T	C	T	C
All students	49	45	50	45	54	47
Gender:						
Male	49	48	52	47	55	48
Female	49	42	48	44	53	47
Ethnicity:						
Hispanic	51	44	53	46	56	48
Black	36	38	36	35	42	41
White/Other	69	69	70	66	69	63
FRL:						
Receives free or reduced-price lunch	43	40	44	41	48	43
Does not receive free or reduced-price lunch	70	65	71	65	73	69
ESE:						
Exceptional education students	22	18	24	21	23	22
Nonexceptional education students	52	48	53	48	57	50
ESOL:						
ESOL levels 1–4	13	12	18	14	21	13
ESOL level 5	54	50	58	54	66	63
Exited over 2 years ago or non-ESOL	54	51	54	50	59	54

iable was used as a classroom covariate since the aggregates of student background variables are sometimes used in multilevel models at higher levels to control for their effects on classrooms.

As different students were in the intervention each year, a separate model was built for each year of the intervention. The analysis for each year included each teacher (and the students assigned to the teacher) for the year, regardless of how many years the teacher participated in the intervention. First, for each year, an initial unconditional two-level model was specified, in which no student-level or teacher-level predictors were included. Using this model, the ICC[1] was computed, which estimated the variability attributed to the teacher level. Next, for each year, the model-building process consisted of two phases: (*a*) a level-1 model expressing the odds of a student being proficient as a function of relevant student predictors and (*b*) a model for the level-2 intercept expressing the odds of students assigned to a teacher being proficient as a function of group predictors, including treatment status. Within each phase, a sequential model-building process was employed, and for each predictor, the fixed effect and heterogeneity in slope coefficient (e.g., variance of the random effect) were examined as suggested by Raudenbush and Bryk (2002). If the unconditional variance of a random effect was not statistically significant, then a model with the relevant effect fixed across teachers was evaluated. The models were examined using the HLM 6.06 software (Raudenbush, Bryk, & Congdon, 2004) with results from the population-average model presented.

The coefficients for the predictors in the model represent the odds of proficiency in logit units. The coefficient and the test for statistical significance for the GROUP variable indicate a difference between the treatment and control groups (research question 1). In addition, the odds ratio (OR) associated for each of the predictor variables was presented. The OR represents the odds of students with a certain characteristic passing the state science assessment compared to students without that characteristic. The OR is a standardized measure of effect that can be used to compare magnitudes across different studies (Lipsey et al., 2012). Rosenthal (1996) describes an odds ratio of 1.5 as a small effect or weak association, 2.5 as a medium effect, and 4.0 as a large effect. The results of both the coefficient and the OR allow us to examine trends across the 3 years (research question 2).

The level-1 or student-level model is expressed as the student-level odds of being proficient as a function of the teacher-level odds of being proficient (β_{0j}) and as a series of predictor variables related to the student. The level-2 or teacher-level model specifies the teacher-level odds of being proficient (β_{0j}) as a function of the individual teacher-level odds of being proficient (γ_{00}), the mean reading score for the teacher, the group variable, and error (μ_{0j}). No teacher-level predictors were entered into the slope parameters (e.g., β_{1j} through β_{8j}). The equation below provides an uncombined equation representing the a priori model.

Level-1 model:

$$\begin{aligned} \mathrm{SCIPROF}_{ij} = {} & \beta_{0j} + \beta_{1j}(\mathrm{GEN}) + \beta_{2j}(\mathrm{BLK}) + \beta_{3j}(\mathrm{WHT}) + \beta_{4j}(\mathrm{FRL}) + \beta_{5j}(\mathrm{ESE}) \\ & + \beta_{6j}(\mathrm{ESOL1-4}) + \beta_{7j}(\mathrm{ESOL5}) + \beta_{8j}(\mathrm{READ}) + r_{ij} \end{aligned}$$

Level-2 model:

$$
\begin{aligned}
\beta_{0j} &= \gamma_{00} + \gamma_{01}(\mathrm{MREAD}) + \gamma_{02}(\mathrm{GROUP}) + u_{0j}\\
\beta_{1j} &= \gamma_{10} + u_{1j}\\
&.\\
&.\\
&.\\
\beta_{8j} &= \gamma_{80} + u_{8j}
\end{aligned}
$$

Results

Descriptive Results

Table 5 presents the percent of treatment and control group students classified as proficient (research question 1). Across all 3 years, the percent of proficient students was greater in the treatment group than the control group, with the difference in the percent proficient being the greatest in Year 3 (T = 54%, C = 47%). Across all 3 years, a higher percent of Hispanics in the treatment group were classified as proficient compared to the percent of Hispanics in the control group (Year 1: T = 51%, C = 44%; Year 2: T = 53%, C = 46%; Year 3: T = 56%, C = 48%). There were some differences in the percent of proficient students classified as ESOL between the treatment and control groups, with the treatment group having a higher percent of proficient students in both ESOL categories (i.e., ESOL levels 1–4 and ESOL level 5) across all 3 years. In the group of students classified as ESOL levels 1–4, in Year 1 there was a 1% difference between the treatment and control groups (T = 13%, C = 12%), while in Year 2 there was a 4% difference (T = 18%, C = 14%), and in Year 3 there was an 8% difference (T = 21%, C = 13%). In the group of students classified as ESOL level 5, there was a 3%–4% difference between the treatment and control groups across all 3 years (Year 1: T = 54%, C = 50%; Year 2: T = 58%, C = 54%; Year 3: T = 66%, C = 63%).

Table 5 also presents the percent of students classified as proficient by demographic groups for each year of the intervention. Across the 3 years, a higher percent of males were proficient compared to the percent of females. For the ethnicity variable, since the Other group made up a very small percent of the sample (about 2%) and performed similarly to Whites, these two groups were combined for all analyses. Among the different ethnic groups, students classified as White/Other had the highest percent of proficient students, and Blacks had the lowest percent. Across the years, students who did not receive free or reduced-price lunch had a higher percent proficient when compared to students who did receive free or reduced-price lunch. Students who were not classified as ESE had a higher percent proficient compared to students who were classified as ESE. Students classified as ESOL level 5 and non-ESOL had a higher percent proficient compared to ESOL levels 1–4. Generally the ESOL level 5 and non-ESOL students had a similar percent proficient, and at some points the ESOL level-5 students actually had a higher percent proficient than non-ESOL students.

Results of Hierarchical Generalized Linear Models

While the descriptive results shown in Table 5 provide initial insight into general trends in the percent of students classified as proficient across the 3 years, the results of the HGLMs provide a more rigorous examination of the odds of being classified as proficient. Results from the unconditional and intermediate models are available upon request. The ICC for Year 1 was computed as .35, for Year 2 as .37, and for Year 3 as .37. Thus, the variance attributed to teachers was about 35% for Year 1 and 37% for Years 2 and 3, indicating a multilevel analysis was appropriate as there was a clustering effect. In addition, the variance of the intercept (μ_{0j}) was statistically significant in each of the models, suggesting a significant amount of variance in students' science proficiency between teachers.

Discussion of final models. The results from the final model for each year are displayed in Table 6. In each of the models, the variance of the random effects associated with most of the predictors was fixed to zero, suggesting that the effects of these variables on the odds of being proficient were the same regardless of which teacher the student was assigned to. For example, the effect of GEN on a student's odds of being proficient in science was the same regardless of the classroom the student was in. It is important to note that the models for the different years were similar in this respect.

The level-1 model results apply to all students in both the treatment and control groups. In the level-1 model for Year 1, GEN, ESE, ESOL1–4, and READ were all statistically significant. Most predictors had an OR below 1, suggesting that the specific demographic group that was coded as "1" was lower in terms of odds of being

Table 6. Results of Final Model for Each Year

	Year 1			Year 2			Year 3		
Fixed Effect	Coefficient	*SE*	OR	Coefficient	*SE*	OR	Coefficient	*SE*	OR
Intercept (γ_{00}):	.35**	.13	1.42	.51**	.14	1.67	.64**	.14	1.89
MREAD (γ_{01})	.06**	.00	1.06	.14**	.01	1.15	.13**	.01	1.14
GROUP (γ_{02})	−.02	.12	.98	.25*	.13	1.28	.31*	.13	1.36
GEN (γ_{10})	−.60**	.07	.55	−.77**	.07	.46	−.78**	.07	.46
BLK (γ_{20})	−.14	.11	.87	−.30**	.11	.74	−.37**	.11	.69
WHT (γ_{30})	.14	.14	1.15	.10	.13	1.10	−.07	.13	.93
FRL(γ_{40})	−.16	.09	.85	−.27**	.09	.77	−.08	.10	.92
ESE (γ_{50})	−.35**	.13	.71	−.13	.13	.88	.00	.16	1.00
ESOL1–4 (γ_{60})	−.56**	.15	.57	−.34**	.13	.71	−.86**	.11	.42
ESOL5(γ_{70})	−.11	.09	.90	.07	.09	1.07	−.05	.09	.95
READ (γ_{80})	.05**	.00	1.05	.10**	.00	1.10	.10**	.00	1.11
Random Effect	Variance	*SD*		Variance	*SD*		Variance	*SD*	
Intercept (u_{0j}):	.37	.61**		.51**	.71		.63**	.79	
ESE (u_{5j})	–	–		–	–		.73[a]	.85	
READ (u_{8j})	.00[a]	.01		.00[a]	.02		.00[a]	.02	

Note.—The effect of GEN, BLK, WHT, FRL, ESOL1–4, and ESOL5 was fixed in all models since the effect of these variables did not differ across teachers, as suggested by a nonsignificant variance component. The effect of ESE was fixed in the Year 1 and Year 2 models since the effect of this variable did not differ across teachers, as suggested by a nonsignificant variance component.

[a] The variance of this variable was not statistically significant, yet fixing the variance across level-2 units led to a statistically significant decrease in model fit.

* $p < .05$.

** $p < .01$.

proficient as compared to the reference group that was coded as "0." The coefficient for GEN was –0.60 ($t[6772] = -8.44$, $p < .001$), indicating that across teachers, an average logit decrease of 0.60 was observed for a female student compared to a male student, after controlling for the effects of all other level-1 variables. The OR of 0.76 indicates that the odds of a female student being proficient were lower by 24% compared to a male student; this is considered a small effect. The coefficient for READ was 0.05 ($t[212] = 33.69$, $p < .001$), indicating that across teachers, for every one standard deviation unit increase in READ, we expect a logit increase of 0.03 in science proficiency. The OR of 1.05 indicates that there was a 5% increase in the odds of being proficient for each one standard deviation increase on READ; this is considered a small effect. It should be noted that GEN, ESE, and ESOL1–4 are statistically significant even after controlling for READ, which is correlated with many of these variables, thus reducing their effects. The coefficient and OR estimate for each of these variables for Year 2 and Year 3 can be interpreted in a similar way as for Year 1.

In Year 1, the coefficient for GROUP was –0.02 ($t[210] = -0.15$, $p = .879$), which was not statistically significant (research question 1). On average, teachers in the treatment group and control group had a similar logit score. The OR of 0.98 indicates that there was a 2% lower odds of being proficient for students of teachers in the treatment group compared to the control group; this is considered a negligible effect. In Year 2, the coefficient for GROUP was 0.25 ($t[201] = 1.94$, $p = .053$), which approached statistical significance (research question 1). On average, teachers in the treatment group had a higher logit score by 0.25. The OR of 1.28 indicates that there was a 28% higher odds of being proficient for students of teachers in the treatment group compared to the control group; this is considered a small but meaningful effect. In Year 3, the coefficient for GROUP was 0.31 ($t[191] = 2.29$, $p = .023$), which was statistically significant (research question 1). On average, teachers in the treatment group had a higher logit score by 0.31. The OR of 1.36 indicates that there was a 36% higher odds of being proficient for students of teachers in the treatment group compared to the control group; this is considered a small but meaningful effect.

Overall, the trend in differences between the treatment and control groups across the 3 years (research question 2) indicates that there was no significant difference in Year 1, the difference approached statistical significance in Year 2 (in which case, it may be more meaningful to examine the effect size), and there was a statistically significant difference in Year 3. Based on the OR estimates, the difference widened a little from Year 2 to Year 3, although the effect size was small in both years.

Discussion, Contributions, and Future Research

The P-SELL intervention in the study was motivated by the need to simultaneously promote science learning and English-language development of ELLs in elementary schools. Specifically, this study examined the effect of the intervention on science proficiency of fifth-grade students measured by high-stakes science assessment over a 3-year intervention.

Discussion

In Year 1, there was no intervention effect detected. In Year 2, the difference in the odds of students assigned to a treatment group teacher being classified as proficient, as compared to students assigned to a control group teacher, approached statistical significance. In Year 3 there was a statistically significant difference in the odds. Since the Year 2 difference approached statistical significance, it is important to examine the effect size, as this may be more important. In both Year 2 and Year 3, the effect size was small and increased slightly from Year 2 to Year 3, indicating a positive effect of the intervention on student achievement in the treatment group (research question 1). According to Lipsey et al. (2012), a small effect size is practically meaningful for an educational intervention that implements a school-based curriculum and uses a standardized test as an outcome measure.

Furthermore, our finding suggests that there was a lag in the impact of the intervention on student achievement (research question 2), which is consistent with other studies indicating a delayed effect of professional development on student outcomes (Kreider & Bouffard, 2006; Silverstein et al., 2009). It may take time for teachers to effectively implement a new curriculum that results in impact on student outcomes. This finding highlights the need to examine how teachers learn to implement or even adapt an intervention over multiple years and to examine what mechanisms or institutional supports are needed for implementation or adaptation over time. This finding also reminds educational practitioners that although they may look for immediate results of an intervention or program, it takes time for an intervention or program to take root and demonstrate effectiveness.

We attribute the effect of the intervention to the three key features of the intervention—standards-based, inquiry-oriented, and ELL-focused. The curriculum materials (i.e., student books, teachers' guide, and supplies) were designed to be educative for teacher learning (Davis & Krajcik, 2005; Davis et al., 2014; Drake et al., 2014), and emphasized these key features throughout. In addition, the teacher workshops followed the core and structural features of effective professional development (Desimone, 2009; Garet et al., 2001) and emphasized these key features. The results do not allow us to tease apart the relative importance of each of these features of the intervention; rather, the results indicate that to address complex educational issues, an intervention needs to pursue multiple goals simultaneously.

The results indicate that an intervention that is targeted at a specific student demographic group, ELLs in our intervention, could benefit all students. The analyses we conducted demonstrated the effect on mean achievement of all students in the sample. In addition, differences existed between the treatment group and control groups in both the ESOL levels 1–4 group and the Hispanic group (see Table 5), although small sample sizes within subgroups and limited variability restricted the analyses that could be conducted (see the discussion of this limitation below).

Contributions

This study makes important contributions to the literature and educational practices. First, although there has been a call for integrating content areas with English language and literacy for ELLs, educational interventions to meet this need have been limited (Buxton & Lee, 2014; Janzen, 2008; Lee, 2005). The curricular and professional development intervention in this study was built on previous research demonstrating the synergy of science and English-language development for ELLs. Particularly, the intervention highlighted the effectiveness of hands-on, inquiry-based science with ELLs, and strategies for English-language development were embedded within science instruction (August et al., 2009; Hampton & Rodriguez, 2001; Lee et al., 2008).

Second, effective science instruction for ELLs should be conceptualized and implemented in the context of standards-based and accountability policies in urban settings (Marx & Harris, 2006; Southerland, Smith, Sowell, & Kittleson, 2007). ELLs should engage in a rigorous science curriculum aligned with reform-oriented practices according to science standards. If high-quality instructional materials that meet current science standards are difficult to find, materials that also take into account the cultural and linguistic diversity of today's classrooms are even scarcer (Lee & Buxton, 2008). In addition, science supplies should be provided for hands-on science, which is particularly effective with ELLs and students with limited experience with formal school science. The results of this study highlight the importance of high-quality instructional materials and supplies to provide effective science instruction for all students, and ELLs in particular.

Finally, professional development interventions should consider that many elementary school teachers are inadequately prepared in knowing the content and processes of science disciplines, using subject-specific teaching strategies, and meeting the learning needs of ELLs (National Center for Education Statistics, 2012). As such, interventions should reach out to those teachers who normally do not volunteer for professional development opportunities (Gamoran et al., 2003; Garet et al., 2001). The results of the study are noteworthy, considering that the intervention involved all grade-level teachers in the participating schools through a school-wide initiative.

Future Research

The results of this study point to directions for future research on educational interventions to promote science learning of ELLs. The current study had several methodological strengths. First, the study used a randomized control trial (RCT) design to test the effect of the intervention involving 31 treatment schools and 32 control schools with approximately 7,000 students in the sample each year. The RCT design has been called for as a rigorous methodological design to test an intervention's effect (McDonald, Keesler, Kauffman, & Schneider, 2006; Wayne et al., 2008). Second, based on random selection of schools, the results can be generalized to other elementary schools in the school district and other similar districts. Third, because the intervention involved all fifth-grade science teachers in the participating schools, rather than a self-selected group of volunteer teachers, the results are more representative of the teaching population. Fourth, the study involved three cohorts of students over the 3 years of the study, which served as replications and allowed examination of trends in science proficiency over time. Finally, as the study was conducted in a state where high-stakes science assessment counted toward school accountability, the results provide insight for other states that include science as part of their state accountability systems.

We also note limitations to our study. First, since the outcome variable was binary, larger samples are needed to detect change and variability (Schochet, 2013). Because the teacher-level variance was not statistically significant, differences between ELLs and non-ELLs in the treatment and control groups were not fully examined using multilevel modeling. The difference between the treatment and control groups in the percent of ESOL levels 1–4 students passing increased every year, favoring the treatment group. However, the sample size of the ESOL levels 1–4 group was much smaller than the entire sample. Similar to the ESOL levels 1–4 group, the difference between the treatment and control groups in the percent of Hispanic students passing was noticeable each year, favoring the treatment group. Although this group had the most ELLs and was most positively affected by the intervention, again the limited variability restricted the analyses.

Second, the study was limited because we had scores available on the high-stakes science assessment administered at the end of fifth grade only. Therefore, we could not control for any preexisting differences in science achievement between the treatment and control groups. Using an RCT, the results of the study provided the causal evidence for the intervention effect. Yet, had there been a pre- and post-measure administered, the study could have provided stronger causal evidence by comparing change scores between the treatment and control groups. Similarly, unlike high-stakes assessment in reading or math administered at grades 3, 4, and 5, science was administered only in fifth grade and thus performance from previous grades could not be used as a predictor variable, which also would have increased the causal evidence.

A logical next step for further research involves testing a change model of how a curricular and professional development intervention affect teachers' knowledge and practices and how teacher change, in turn, impacts student outcomes (Desimone, 2009). Creating such a change model based on an effective intervention could help identify critical features of professional development that result in teacher change and student achievement. Using rigorous research designs will enable researchers to test the efficacy and effectiveness of educational innovations. In turn, this increasing knowledge base will allow educators to respond to the urgent call to simultaneously promote science learning and English-language development among the growing population of ELLs in the current policy context of high-stakes assessment and accountability.

Notes

Jaime Maerten-Rivera works in the area of education assessment at the State University of New York at Buffalo. Soyeon Ahn is associate professor of the Research, Measurement, and Evaluation program, School of Education and Human Development, University of Miami. Okhee Lee is professor of childhood education in the Steinhardt School of Culture, Education, and Human Development at New York University. Kimberly Lanier is the director of the Faculty Institute of Teaching and Learning at Miami Dade College. Jennifer Diaz is the director of education at the Everglades Foundation, Inc. Correspondence may be sent to Okhee Lee at olee@nyu.edu. The authors acknowledge Jane Sinagub for her editorial feedback on a draft version of the article. The authors give special thanks to all of the teachers who participated in the research. This research was supported by the Institute of Education Sciences, U.S. Department of Education, through grant no. R305A090281 to the University of Miami. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

1. Congruent with Snijders & Bosker (1999), the intraclass coefficient for the school level was determined from the following formula: $\tau_{00}/(\tau_{00} + 3.29)$, where τ_{00} is the estimated level-2 variance for the model.

References

Abell, S. (2007). Research on science teachers' knowledge. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 1105–1149). Mahwah, NJ: Erlbaum.

August, D., Branum-Martin, L., Cardenas-Hagan, E., & Francis, D. (2009). The impact of an instructional intervention on the science and language learning of middle grade English language learners. *Journal of Research on Educational Effectiveness*, **2**, 345–376.

Blanchard, M. R., Southerland, S. A., & Granger, E. M. (2009). No silver bullet for inquiry: Making sense of teacher change following an inquiry-based research experience for teachers. *Science Education*, **93**(2), 322–360.

Buxton, C. A., & Lee, O. (2014). English language learners in science education. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research in science education* (2nd ed., pp. 204–222). Mahwah, NJ: Erlbaum.

Davis, E., & Krajcik, J. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, **34**, 3–14.

Davis, E. A., Palincsar, A. S., Arias, A., Bismack, A., Marulis, L., & Iwashyna, S. (2014). Designing educative curriculum materials: A theoretically and empirically driven process. *Harvard Educational Review*, **84**(1), 24–52.

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

Diamond, B. S., Maerten-Rivera, J., Rohrer, R. E., & Lee, O. (2014). Effectiveness of curricular and professional development intervention on elementary teachers' science content knowledge and student achievement outcomes: Year 1 results. *Journal of Research in Science Teaching*, **51**, 635–658.

Drake, C., Land, T. J., & Tyminski, A. M. (2014). Using educative curriculum materials to support the development of prospective teachers' knowledge. *Educational Researcher*, **43**(3), 154–162.

Gamoran, A., Anderson, C. W., Quiroz, P. A., Secada, W. G., Williams, T., & Ashmann, S. (2003). *Transforming teaching in math and science: How schools and districts can support change*. New York: Teachers College.

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

Goldenberg, C. (2013, Summer). Unlocking the research on English learners: What we know—and don't yet know—about effective instruction. *American Educator*, 4–11, 38.

González, N., Moll, L. C., & Amanti, C. (2005). *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Mahwah, NJ: Erlbaum.

Hampton, E., & Rodriguez, R. (2001). Inquiry science in bilingual classrooms. *Bilingual Research Journal*, **25**, 461–478.

Heller, J. I., Daeler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, **49**(3), 333–362.

Janzen, J. (2008). Teaching English language learners in the content areas. *Review of Educational Research*, **78**, 1010–1038.

Kennedy, M. M. (1998). Education reform and subject matter knowledge. *Journal of Research and Science Teaching*, **35**, 249–263.

Kreider, H., & Bouffard, S. (2006). Questions and answers: A conversation with Thomas R. Guskey. *The Evaluation Exchange*, **11**(4). Retrieved from http://www.gse.harvard.edu/hfrp/eval/issue32/qanda.htm

Lee, O. (2005). Science education and English language learners: Synthesis and research agenda. *Review of Educational Research*, **75**, 491–530.

Lee, O., & Buxton, C. A. (2008). Science curriculum and student diversity: Culture, language, and socioeconomic status. *Elementary School Journal*, **109**, 123–137.

Lee, O., Deaktor, R., Enders, C., & Lambert, J. (2008). Impact of a multi-year professional development intervention on science achievement of culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching*, **45**(6), 726–747.

Lee, O., Maerten-Rivera, J., Penfield, R. D., LeRoy, K., & Secada, W. G. (2008). Science achievement of English language learners in urban elementary schools: Results of a first-year professional development intervention. *Journal of Research in Science Teaching*, **45**(1), 31–52.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms*. Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncser/

Maerten-Rivera, J., Myers, N., Lee, O., & Penfield, R. (2010). Student and school predictors of high-stakes assessment in science. *Science Education*, **94**(6), 937–962.

Marx, R. W., & Harris, C. J. (2006). No Child Left Behind and science education: Opportunities, challenges, and risks. *Elementary School Journal*, **106**, 467–477.

McDonald, S.-K., Keesler, V. A., Kauffman, N. J., & Schneider, B. (2006). Scaling-up exemplary interventions. *Educational Researcher*, **35**, 15–24.

National Center for Education Statistics. (2012). *The condition of education 2012*. Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from http://nces.ed.gov/programs/coe/indicator_ell.asp

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

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

NRC (National Research Council). (2007). *Taking science to school: Learning and teaching science in grades K–8*. Washington, DC: National Academy.

NRC (National Research Council). (2011). *A framework for K–12 science education: Practices, crosscutting themes, and core ideas*. Washington, DC: National Academy.

Penuel, W. R., Fishman, B., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, **44**(4), 921–958.

Raudenbush, S., Bryk, A., & Congdon, R. (2004). *HLM™ 6 for Windows*. Lincolnwood, IL: Scientific Software International.

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

Remillard, J. T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, **75**(2), 211–246.

Rosebery, A. S., & Warren, B. (Eds.). (2008). *Teaching science to English language learners: Building on students' strengths*. Arlington, VA: National Science Teachers Association.

Rosenthal, J. A. (1996). Qualitative descriptors of strength of association and effect size. *Journal of Social Service Research*, **21**(4), 37–59.

Schochet, P. (2013). Statistical power for school-based RCTs with binary outcomes. *Journal of Research on Educational Effectiveness*, **6**(3), 263–294.

Silverstein, S. C., Dubner, J., Jon, M., Glied, S., & Loike, J. D. (2009). Teachers' participation in research programs improves their students' achievement in science. *Science*, **326**(5951), 440–442.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, **38**, 553–573.

Southerland, S., Smith, L. K., Sowell, S. P., & Kittleson, J. M. (2007). Resisting unlearning: Understanding science education's response to the United States' national accountability movement. In L. Parker (Ed.), *Review of research in education* (Vol. **31**, pp. 45–77). Washington, DC: American Educational Research Association.

TESOL (Teachers of English to Speakers of Other Languages). (2006). *PreK–12 English language proficiency standards*. Alexandria, VA: Author.

van Driel, J. H., Berry, A., & Meirink, J. (2014). Research on science teacher knowledge. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of research in science education* (2nd ed., pp. 848–870). Mahwah, NJ: Erlbaum.

Warren, B., Ballenger, C., Ogonowski, M., Rosebery, A., & Hudicourt-Barnes, J. (2001). Rethinking diversity in learning science: The logic of everyday language. *Journal of Research in Science Teaching*, **38**(5), 529–552.

Wayne, A. J., Yoon, K. S., Zhu, P., Cronen, S., & Garet, M. S. (2008). Experimenting with teacher professional development: Motives and methods. *Educational Researcher*, **37**, 469–479.

Weaver, D., & Dick, T. (2009). Oregon mathematics leadership institute project: Evaluation results on teacher content knowledge, implementation fidelity, and student achievement. *Journal of Mathematics and Science: Collaborative Explorations*, **11**, 57–84.

Wilson, C. D., Taylor, J. A., Kowalski, S. M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, **47**(3), 276–301.