J Sci Teacher Educ
DOI 10.1007/s10972-013-9376-6

ELEMENTARY SCIENCE TEACHER EDUCATION

Improving Science and Literacy Learning for English Language Learners: Evidence from a Pre-service Teacher Preparation Intervention

Jerome M. Shaw · Edward G. Lyon · Trish Stoddart · Eduardo Mosqueda · Preetha Menon

Abstract This paper present findings from a pre-service teacher development project that prepared novice teachers to promote English language and literacy development with inquiry-based science through a modified elementary science methods course and professional development for cooperating teachers. To study the project's impact on student learning, we administered a pre and post assessment to students (N = 191) of nine first year elementary teachers (grades 3 through 6) who experienced the intervention and who taught a common science unit. Preliminary results indicate that (1) student learning improved across all categories (science concepts, writing, and vocabulary)—although the effect varied by category, and (2) English Language Learner (ELL) learning gains were on par with non-ELLs, with differences across proficiency levels for vocabulary gain scores. These results warrant further analyses to understand the extent to which the intervention improved teacher practice and student learning. This study confirms the findings of previous research that the integration of science language and literacy practices can improve ELL achievement in science concepts, writing and vocabulary. In addition, the study indicates that it is possible to begin to link the practices taught in pre-

J. M. Shaw (✉) · T. Stoddart · E. Mosqueda · P. Menon
Department of Education, University of California, 1156 High Street, Santa Cruz, CA 95064, USA
e-mail: jmlshaw@ucsc.edu

T. Stoddart
e-mail: stoddart@ucsc.edu

E. Mosqueda
e-mail: mosqueda@ucsc.edu

P. Menon
e-mail: pmenon@ucsc.edu

E. G. Lyon
Mary Lou Fulton Teachers College, Arizona State University, 1050 S. Forest Mall,
Tempe, AZ 85287-5411, USA
e-mail: eglyon@asu.edu

Published online: 05 January 2014

service teacher preparation to novice teacher practice and student learning outcomes.

Keywords Student achievement · English Language Learners · Intervention study · Pre-service elementary teachers · Science and literacy integration

Introduction

Two critical challenges face science education today: (1) improving the teaching and learning of students who do not speak English as a first language (English Language Learners, or ELLs) and (2) the preparation of the teachers who serve them. The population of ELLs is rapidly expanding across the United States; it is projected that one in every four students in the U.S. will speak English as a second language by 2025 [National Center for Education Statistics (NCES), 2011; National Clearing House for English Language Acquisition, 2007; U.S. Census Bureau, 2010]. For at least 30 years, ELLs' achievement in science, language, and literacy has lagged behind that of native English speakers (Buxton, 2006; Grigg, Daane, Jin, & Campbell, 2002; Lee & Luykx, 2006; NCES, 2010). For example, on the 2009 National Assessment of Educational Progress (NAEP), the average 4th grade scaled science score was 114 (SE = .8) for ELLs, compared to 154 (SE = .2) for non-ELLs (NCES, 2011). ELLs are also less likely to pursue advanced degrees in science (Commission on Professionals in Science and Technology, 2007; National Academy of Sciences, 2010) or to perceive science subjects as relevant to their lives (Aikenhead, 2006; Buxton, 2006; Barton, 2003; Lynch, 2001; Rodriguez, 1998).

An emerging body of research, however, has demonstrated that integrating the development of English language and literacy with contextualized science inquiry improves ELLs' achievement in science (Bravo & García, 2004; Cervetti, Pearson, Barber, Hiebert, & Bravo, 2007; Lee, Maerten-Rivera, Penfield, LeRoy, & Secada, 2008; Ovando & Combs, 2012; Rivet & Krajcik, 2008; Rosebery & Warren, 2008). Many teaching practices promoted through this research, such as collaborative inquiry, challenging science content, and modeling/practice with complex language forms, echo principles from *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* [National Research Council (NRC), 2012] as well as the Common Core State Standards (2010). As the *Framework* recognizes, scientific inquiry involves more than hands-on activities; it also involves active thinking and discourse around such activities (NRC, 2012). During inquiry, language and science content learning intersect as students construct oral and written explanations and engage in argumentation from evidence (Lee, Quinn & Valdés, 2013; Santos, Darling-Hammond, & Cheuk, 2012).

Thus, the first critical challenge—*improving* science teaching and learning for ELLs—can be addressed by instruction that integrates contextualized science inquiry with English language and literacy development. The second challenge—*preparing* teachers to effectively instruct ELLs—can be addressed in teacher education programs as well as through professional development for practicing classroom teachers. We focus on the second challenge through our work with pre-service teachers, using improvement in student learning as the outcome measure.

Despite the severity and persistence of the achievement gap between ELLs and their native English-speaking peers, few teachers receive education in how to teach subject content to ELLs (Ballantyne, Sanderman, & Levy, 2008; Darling-Hammond, 2006; Gándara, Maxwell-Jolly, & Driscoll, 2005; Villegas & Lucas, 2002). It is not surprising that most teachers do not feel prepared to teach ELLs (Parsad, Lewis, Farris, & Greene, 2001). In a recent national survey, only 15 % of elementary teachers—the focus group for our study—reported feeling well prepared to teach science to ELLs (Banilower et al., 2013). The consequence is that ELLs are the group least likely to have a qualified or experienced science teacher (Business Higher Education Forum, 2006; California Council on Science and Technology, 2010; Oakes, Joseph, & Muir, 2004).

Researchers have studied an array of approaches and strategies to help pre-service teachers become more prepared to teach content to culturally and linguistically diverse students, including: confidence and knowledge building through a community-based experience in a non-formal educational setting (Buck & Cordes, 2005), equitable instructional and assessment strategies in mainstream classrooms (Bianchini, Johnston, Oram, & Cavazos, 2003; Lyon, 2013), and the integration of multicultural strategies in science instructional units (Suriel & Atwater, 2012). However, the studies cited above focus primarily on teacher outcomes (e.g., teacher knowledge, beliefs, and practices), instead of how the intervention ultimately impacts *student* outcomes. Intervention studies that include student outcomes typically address the impact of professional development of in-service teachers, not pre-service preparation (cf. Hart & Lee, 2003; Lee et al., 2008). Thus, for pre-service teachers, the second challenge remains largely unresolved with respect to the number of educators reached and the inclusion of approaches that (a) *integrate* science and English language and literacy instruction, and (b) are supported with student learning data. In this paper, we argue that the project (described next) offers a viable means to address both challenges with evidentiary support of the project's promise for impacting student learning.

The ESTELL Project

The Effective Science Teaching for English Language Learners (ESTELL) project focused on preparing pre-service teachers to integrate the teaching of science with English language and literacy for ELLs. Education faculty and science education teachers from four universities collaborated on the project's (a) instructional framework and (b) intervention—each described below.

ESTELL Instructional Framework

The ESTELL project's instructional framework is based on research exploring teaching practices that promote science, language, and literacy learning, including: (a) the U.S. Department of Education funded *Center for Research on Education*

Diversity and Excellence (CREDE) project (Doherty & Pinal, 2004; Hilberg, Tharp, & DeGeest, 2000) and (b) a set of *NSF funded science-language-literacy integration projects* (Cervetti et al., 2007; Lee et al., 2008; Stoddart, Pinal, Latzke, & Canaday, 2002). Both approaches identified a common set of specific and observable teacher actions that a substantial body of empirical research has demonstrated raise the achievement of culturally and linguistically diverse students and improves their motivation to learn. For the instructional framework (Stoddart, Solis, Tolbert, & Bravo, 2010), these six teaching practices are: (1) facilitating collaborative inquiry, (2) promoting science talk, (3) literacy in science, (4) scaffolding and development of language in science, (5) contextualizing science activity, and (6) promoting complex thinking (see Table 1). Although the integration of these practices into science teaching has been shown to improve the learning of ELLs with experienced teachers, the challenge is to prepare new teachers to actually use this integrated pedagogy with ELLs, and to evaluate the impact on their students' learning of science and literacy.

ESTELL Pre-service Teacher Education Intervention

The ESTELL project designed, implemented, and evaluated a pre-service science teacher education intervention based on the above instructional framework and teacher education research demonstrating that novice teachers need to (a) observe and experience explicit models of the pedagogy they are learning to teach (Abell & Cennamo, 2004; Goldman, Pea, Baron, & Derry, 2007; Roth et al., 2011) and (b) practice instructional approaches with the student population they are being prepared to teach with intensive feedback, coaching, and support (Joyce & Showers, 1995; Loucks-Horsley, Hewson, Love, & Stiles, 1998; Speck & Knipe, 2001). Accordingly, key elements of the intervention are (a) a restructured elementary *science methods course* used in post-baccalaureate teacher education programs, and (b) *professional development* for practicing teachers who would be mentoring the participants. The components of the ESTELL pre-service teacher education program are presented in Fig. 1.

Revised Science Methods Course

A team of four project science method instructors developed and modeled five science units (in the physical, life, and earth sciences) that explicitly reflect the project's instructional practices in their elementary science methods course.[1] While each lesson illustrated all of the instructional practices, individual lessons highlighted one or two of the practices to make it easier for student teachers to engage with the project framework. In concert with the methods course, all project pre-service teachers completed a 15-week student teaching practicum in a K-6 classroom, during which they had increasing responsibility for classroom instruction. The pre-service teachers used project lesson plan templates to design and implement science lesson activities during their student teaching field experiences.

[1] Units are available at http://j.mp/RBUq5O.

Table 1 ESTELL project instructional framework

Instructional practice	Description
Facilitating collaborative inquiry	Promoting effective learning communities through inclusive and collaborative student engagement that leads to a joint learning product in science activity. Through collaboration, students recognize the social construction of scientific knowledge
Promoting science talk	Promoting, scaffolding, and providing feedback on student use of academic science discourse and sustained dialogue about science ideas both between student-teacher and between students
Literacy in science	Providing students with opportunities for written or verbal language expression that is authentic in nature and facilitated with literacy tools (e.g., science notebooks). Teacher uses curriculum-based science inquiry and content terms, and provides students with opportunities to use these words
Scaffolding and development of language in science	Providing opportunities for English language development through meta-linguistic awareness and modified/comprehensible teacher talk as well as providing scaffolds to amplify the science content of ELLs, such as paralinguistic cues, multi-sensory experiences, and visual representations
Contextualizing science activity	Eliciting and incorporating students' knowledge about personal-home-community experiences and local and global environment/ecology in science activity
Promoting complex thinking	Promoting, scaffolding, and providing feedback on students' scientific reasoning, understanding of key and complex science concepts, and development of inquiry skills

Professional Development for Cooperating Teachers and Teacher Supervisors

Pre-service teachers were placed in the classrooms of CTs and observed periodically by a supervisor. Both cooperating teacher (CTs) and teacher supervisors (TSs) had attended a 2-day professional development institute led by two project members with extensive professional development experience. During the professional development, CTs and TSs (a) learned about the instructional framework, (b) analyzed instructional exemplars reflecting the framework's practices, and (c) received coaching in mentoring techniques and in using a project-developed observation rubric—all so that they could provide effective coaching, mentoring, and feedback to support the project pre-service teachers in implementing the framework's practices during student teaching.

Analysis of pre-service teacher survey and classroom observation data demonstrated that Project-trained pre-service teachers had significantly more knowledge about effective teaching practices for ELLs and used these practices more frequently in their student teaching practicum than a comparison group of pre-service teachers in a 'business as usual' pre-service teacher education program (Stoddart, 2013; Stoddart & Mosqueda, in press). The analysis reported in this paper investigates the impact of the instructional practice of first year teacher Project graduates on the science, language and literacy learning of 3rd through 6th grade students in their classrooms.

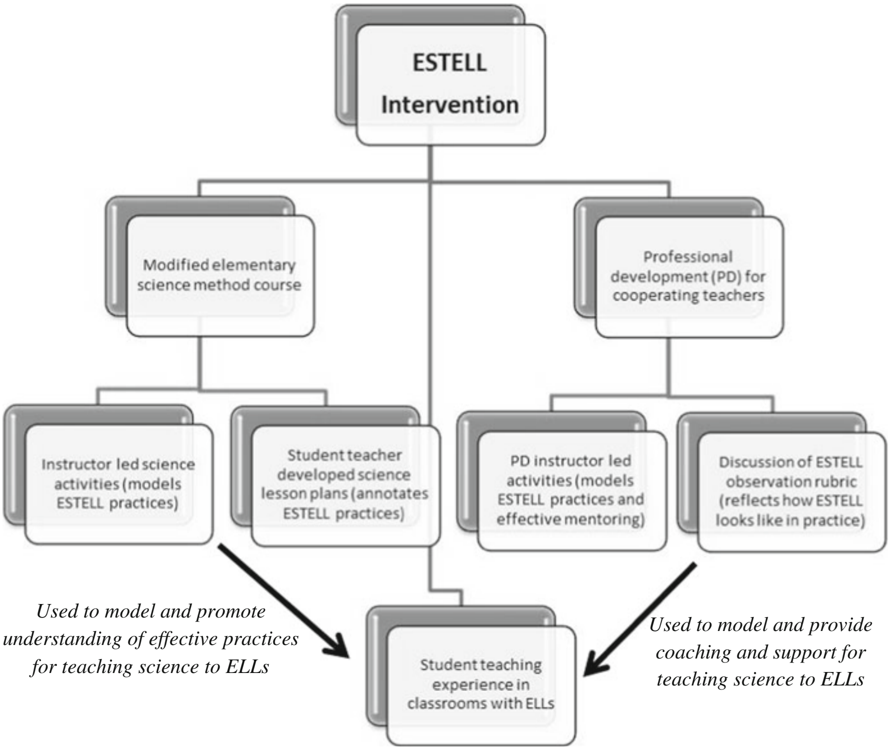

Fig. 1 ESTELL pre-service teacher education program model

Research Methods

The study was carried out in classrooms of nine first year elementary teachers (FYTs), who all participated in the project intervention. The FYTs all taught the same instructional science unit using a common curriculum. Each of the teachers adapted the curriculum unit to include the project's teaching practices as they deemed appropriate. Pre- and post-student achievement data (Science Concepts, Science Writing, Science Vocabulary) were collected on 191 3rd through 6th grade students in the nine FYT classrooms. The sample, curriculum, and assessment are described later.

Research Questions

Three primary research questions guided the analysis of the student achievement data:

1. Do the 3rd through 6th grade students of project-trained first year teachers demonstrate gains in science achievement (as measured by a project-developed assessment) from before to after a single science unit? *If so…*
2. Did these gains vary by achievement category?

3. Did these gains vary by the students' level of English language proficiency?

Teacher Sample

Nine first year teachers (FYTs) were recruited from the pool of project pre-service teacher participants in years 2 and 3 of the project. FYT selection was based upon the following *required* criteria: (1) participation in the project (enrolled in the revised science methods course and completed a pre/post survey), and (2) mentored by a CT who participated in the project's professional development during the time the FYT engaged in her/his teaching practicum. Further *desired* characteristics included (1) being a *full* time teacher in a grade 3–6 classroom, (2) having a minimum of 25 % ELL students, and (3) being employed in the focal state. Our selection process resulted in nine FYTs who, having met all of the required criteria and most of the desired characteristics, agreed to participate in the study. Of these nine, two FYTs taught 3rd grade, three taught 4th grade, three taught 5th grade, and one taught 6th grade. All nine were females between the ages of 20 and 25 and received their credential from the same project-affiliated university. Two FYTs were non-White and three had some second language proficiency. Only one FYT majored in a science-related field. FYT demographic information is presented in Table 2.

Student Sample

A total of 191 students taught by the nine FYTs participated in the study. This number represents those students who completed both pre and post versions of the project assessment (described below) for at least one of the three achievement categories (Science Concepts, Science Writing, Science Vocabulary), and for whom we have ELL status information, including level of proficiency in English as measured by the focal state's English Language Development Test. Former (or re-designated) ELL students were omitted from our analyses to avoid sample bias. Prior studies have shown little understood variances in test functioning with and academic achievement of this student subgroup (de Jong, 2004; Young et al., 2008). Observed over-performance of re-designated ELLs may be due to the additional language support received by these students prior to reclassification (Lee et al., 2006; Shaw, 2009). Students were predominantly Latino/Hispanic (62 %) and 51 % of the students were female. Forty-eight percent ($N = 92$) of the students were ELLs, distributed across four levels of English language proficiency (see Table 3).

Common Curriculum

To make fair comparisons of student achievement across all teachers, each teacher was required to teach *Terrarium Habitats*, a science unit developed by the Great Explorations in Mathematics and Science, commonly known as GEMS, program at the Lawrence Hall of Science (Hosoume & Barber, 1999). *Terrarium Habitats*,

Table 2 First year teacher information

Teacher ID	Cohort	Sex	Race/ ethnicity	2nd language proficiency	Undergraduate field	Grade level
3080	2011	F	White	None	Education	5th
3084	2011	F	White	None	Education	4th
3090	2011	F	White	None	Social Science	5th
3095	2011	F	White	None	Professional program	3rd
4033	2012	F	White	None	Professional program	3rd
4034	2012	F	White	Italian (intermediate)	Social Science	4th
4035	2012	F	White	None	Natural or Physical Science	6th
4055	2012	F	Latino	Spanish (intermediate)	Education	5th
4041	2012	F	Asian	Spanish (intermediate)	Education	4th

Table 3 Student demographics for the total sample (N = 191)

	Total students	Male	Race/ethnicity					English proficiency			
			White	Hispanic/ Latino	Black	Asian	Multi-racial	B/IE	I	EA/A	EO
N	191	94	43	118	11	10	9	22	30	40	99
%	100	49	23	62	6	5	5	12	16	21	52

B Beginning, *EI* Early Intermediate, *I* Intermediate, *EA* Early Advanced, *A* Advanced, *EO* English Only

intended for grades K-6 (with suggested supplemental activities for the younger and older grades), introduces students to key ecological concepts—such as the role of soil, an organism's habitat, and decomposition in an ecosystem, as well as the concept of adaptation—as students build a model terrarium and add to it organisms such as earthworms and isopods. Depending on the frequency and duration of science instruction, teaching of the unit's five investigations can span from one to several weeks.

We selected *Terrarium Habitats* since it was aligned with the focal state's grade 3–6 standards (the grade range of our teachers) and did *not* have an explicit focus on integrating literacy and science. The latter attribute allowed researchers to see the impact of teachers' application of what they learned in their science methods course, and not the impact of the curriculum itself, on student achievement. Other considerations included the demands on teacher time for preparation and implementation, availability of validated corresponding assessment items, and affordability of the materials.

Before teaching *Terrarium Habitats*, all FYTs participated in a 2-day (11-h) workshop that (1) oriented them to the study goals, responsibilities, and logistics; (2) familiarized them with the focal science unit and assessment; (3) enhanced their understanding of the focal science unit's content; and (4) supported them in developing draft plans for teaching the focal science unit.

The ESTELL Student Assessment

Project researchers designed the ESTELL Student Assessment to measure student achievement of curriculum-based learning addressed in the *Terrarium Habitats* unit. The assessment is composed of items from the *Soil Habitats* summative assessment packet (Lawrence Hall of Science, 2007). Those items were developed by non-project affiliated researchers to accompany the *Soil Habitats* curricular unit, which itself is an updated version of *Terrarium Habitats* (which does not have an assessment package). *Soil Habitat*'s assessment items were selected because of their alignment with the Terrarium Habitats content and their inclusion of literacy tasks, such as for writing and vocabulary. Minor revisions were made to some of the *Soil Habitats* assessment items to maintain alignment with the *Terrarium Habitats* curriculum. For example, items that contained the phrase "worm box" had those words changed to "terrarium habitat" because worm boxes were part of the *Soil Habitats* unit but not the *Terrarium Habitats* unit. Such changes did not alter the content focus of the effected items.

The ESTELL Student Assessment consists of three achievement categories (Science Concepts, Science Writing, and Science Vocabulary) and one affective category (Science Attitudes). These categories were combined into a two-part test: Part A includes science writing and attitudes while Part B includes science concepts and vocabulary. In this study, we focus solely on the three achievement categories. Information regarding score ranges and item reliability for those categories is provided in Table 4.

Science Concepts

This category is composed of 14 items (12 multiple-choice; 2 constructed response) across two conceptual clusters: "Decomposition" and "Adaptation." Multiple-choice items ask questions such as "*An earthworm's tail can break off and then grow back. Which of these explains why that is an important adaptation for earthworms?*" The two constructed response items are as follows: "*What could you observe in a terrarium habitat that would be evidence of decomposition?*" (Decomposition cluster) and "*Choose one of these organisms: earthworm, pill bug, or sow bug. For this organism, describe something it does or something about its body that helps protect the organism from predators*" (Adaptation cluster). Due to low internal consistency (based upon Cronbach's alpha) for each conceptual cluster, all concept multiple-choice items were grouped together to form one scale with an internal consistency of .744 for the pre-test and .690 for the post-test (both years combined). George and Mallery (2003) provide the following guidelines for interpreting alpha coefficients: below .5 = unacceptable, .5–.6 = poor, .6–.7 = questionable, .7–.8 = acceptable, .8–.9 = good, above .9 = excellent.

Table 4 Reliability analysis of items

Category	Scale (score range)	Pretest internal consistency	Posttest internal consistency
Multiple choice items			
Science Content	Multiple-choice (0–12)	.744	.690
Science Vocabulary	Inquiry (0–15)	.769	.782
	Content (0–15)	.839	.722
Category	**Scale (score range)**	**Year 1 inter-rater agreement**	**Year 2 inter-rater agreement**
Constructed-response items			
Science Content	Decomposition: constructed-response (0–2)	.48	.38
	Adaptation: constructed-response (0–2)	.44	.45
Science Writing	Argument (0–4)	.41	.36
	Clarity (0–4)	.44	.42
	Use of vocabulary (0–4)	.65	.59

Science Writing

This item consists of the following prompt: *Imagine that you are going to make a place for an earthworm to live* (*in other words, a terrarium habitat*). *What 3 things* (*living or non-living*) *would you put in your habitat to help an earthworm survive? For each thing, explain how that thing would help an earthworm survive.* Students were given written guidelines to (1) *Use complete sentences;* (2) *Use as many science words as possible;* and (3) *Be as clear as possible*. Before writing their response, students are given a 10 min planning period to (1) write words they think you will use (in a word box) and (2) make drawings that they could write about (in a drawing box).

Science Vocabulary

This category consists of 30 multiple-choice items in which students either choose the term that best matches the given definition (e.g., *An organism that breaks down dead plants and animals*) or the term that would best complete a sentence (e.g., *Earthworms are _ that break down dead plants and animals in the soil*). Fifteen of the terms relate to science content vocabulary learned throughout the unit (e.g., decompose, nutrients) while the other 15 relate to more generic science inquiry vocabulary (e.g., predict, observe). The pre-test internal consistency for each scale was as followed: Inquiry vocabulary—.769 (pretest) and .782 (posttest), Content vocabulary—.839 (pretest) and .722 (posttest). Thus, there was acceptable to good internal consistency for these scales.

Scoring

An answer key specifies the single correct response to each multiple-choice item (score range 0–1). To accurately, fairly, and reliably score the three constructed-response items (two concept items and the writing prompt), researchers refined rubrics accompanying the *Soil Habitats* summative assessment package by iteratively reading sample responses, scoring with the revised rubric, and making further refinements to the rubric. Both science concept rubrics had the following 0–4 scale: 0 = Off Topic/No Knowledge/No Response; 1 = Inappropriate Understanding; 2 = Incomplete Understanding; 3 = Contains Essential Features of Curriculum-based Understanding; 4 = Comprehensive Articulation of Curriculum-based Understanding. For the writing prompt, instead of the holistic rubric developed for the science concept items, we developed three analytical rubrics, one each for (1) scientific argument, (2) clarity, and (3) use of science vocabulary. Each of these dimensions is scored on a 0–4 point scale. At least two project researchers scored each response to the constructed-response items. Expert scorers discussed any disagreements to reach a single consensus score. Prior to determining consensus scores, inter-rater agreement using Cohen's kappa ranged from .38 (year 2 decomposition) to .65 (year 1 use of vocabulary). Landis and Koch (1977) suggest the following guidelines for interpreting the kappa coefficient: 0 = poor, .01–.20 = slight, .21–.40 = fair, .41–.60 = moderate, .61–.80 = substantial, and .81–1 = almost perfect (see Table 4).

Data Collection

The ESTELL Student Assessment was administered to students in the FYTs' classrooms prior to and after completing the Terrarium Habitats unit. To reduce the likelihood of student overload, each administration (pre and post) occurred over 2 days, with Part A (administered by FYTs themselves) on the first day and Part B (administered by project researchers in the presence of the FYT) on the second day. Each part took approximately 1 h. All items were read aloud to students as they marked or wrote answers directly on the given assessment packet. Administration was uniform for all students and untimed.

Data Analysis

Analyses are based on pre, post, and gain (*post—pre*) measures of the various achievement categories. The Science Concepts measure (ranging from 0 to 20) for each student was calculated by summing together scores for all 12 multiple choice concept items plus both constructed-response concept items (4 points each). The Science Writing measure (ranging from 0–12) was calculated by summing together scores for all three dimensions (Argument, Clarity, Vocabulary). The Science Vocabulary measure (ranging from 0 to 30) was calculated by summing together scores for all 30 multiple choice vocabulary items (further sub-divided into scores for the 15 inquiry vocabulary items and 15 content vocabulary items). A composite score (ranging from 0 to 62) was created by summing scores on the preceding measures.

In consideration of our exploratory research questions, descriptive statistics were generated to look for general patterns across each measure for the entire student sample as well as patterns in scores by grade level and English language proficiency. English language proficiency levels were based upon focal state's English Language Development Test (ELDT) scores, which place students at one of five levels: Beginning, Early Intermediate, Intermediate, Early Advanced, and Advanced. We used teacher self-report with the same levels when ELDT scores were not available. To produce sufficient and comparable sample sizes for each English language proficiency group, we collapsed Beginning with Early Intermediate levels and Early Advanced and Advanced levels, thus forming three groups of ELL students.

Paired-sample *t*-tests and effect sizes[2] (using Cohen's d statistic, see Footnote 1) were calculated to compare pre to post scores for each measure. In addition, three one-way ANOVAs were calculated to test for statistical differences in the Science Concepts, Science Writing, and Science Vocabulary gains among students of the various English language proficiencies. Composite scores were omitted for these analyses due to low sample size in each English proficiency level (since fewer students completed all three pre and post categories). We do not include grade level analyses given the uneven and limited number of teachers per grade.

Results

We present our findings organized around the three research questions: overall gains, gains by achievement category, and gains by English language proficiency level. Overall descriptive and inferential statistics are displayed in Table 5.

Did the Students of Project Trained First Year Teachers Demonstrate Gains in Science, Language and Literacy Achievement (as Measured by a Project-Developed Assessment) from Before to After a Single Science Unit?

An examination of the composite scores shows that students did demonstrate achievement gains from pre to post. For all students, the mean composite pretest score was 35.07 out of a possible 62 (SD = 11.82), while the posttest score was 41.25 (SD = 9.49), yielding an average gain of 6.18 points (SD = 7.55). This gain was statistically significant ($p < .001$, t = 9.08) with a corresponding effect size of .277 (see Footnote 1). According to Cohen (1988), effect sizes below .2 can be interpreted generally as reflecting no effect; while effect sizes between .2 and .5 reflect a small effect, between .5 and .8 a medium effect, and above .8 a large effect. Thus, the composite gain score reflects a small effect.

Did These Gains Vary by Achievement Category (Science Concepts, Science Writing, Science Vocabulary)?

Gains did vary by achievement category (see Table 5). Out of a possible 20 points, the mean *Science Concepts* gain score was 3.34 (SD = 2.74). Out of 12 points,

[2] Effect size calculated with the following website: http://www.uccs.edu/~lbecker/.

Table 5 Achievement gains for total sample by category

	Possible range	N	Pre *M* (SD)	Post *M* (SD)	Gain *M* (SD)	*T* statistic	*p* value	Effect size (*d*)
Composite	0–62	123	35.07 (11.82)	41.25 (9.49)	6.18 (7.55)	9.08	.000***	.277
Concepts	0–20	128	10.54 (4.06)	13.88 (2.88)	3.34 (2.74)	13.81	.000***	.949
Multiple-choice	0–12	129	6.50 (2.92)	8.85 (2.26)	2.35 (1.93)	13.84	.000***	.900
Constructed-response	0–8	153	3.89 (1.73)	5.02 (1.25)	1.13 (1.82)	7.67	.000***	.749
Writing	0–12	176	4.57 (2.17)	6.06 (2.10)	1.49 (2.34)	8.44	.000***	.698
Argument	0–4	174	1.26 (1.02)	1.76 (1.01)	.500 (1.15)	5.75	.000***	.493
Clarity	0–4	174	1.63 (.814)	1.94 (.727)	.316 (.803)	5.19	.000***	.402
Vocabulary	0–4	174	1.66 (.719)	2.43 (.901)	.770 (1.04)	9.78	.000***	.945
Vocabulary	0–30	167	18.94 (6.74)	20.26 (6.78)	1.32 (5.60)	3.04	.000***	.195
Content	0–15	167	9.40 (3.38)	10.46 (3.39)	1.07 (2.84)	4.88	.003**	.316
Inquiry	0–15	167	9.55 (3.77)	10.71 (3.65)	1.16 (2.88)	5.18	.000***	.313

* $p < .05$ ** $p < .01$ *** $p < .001$

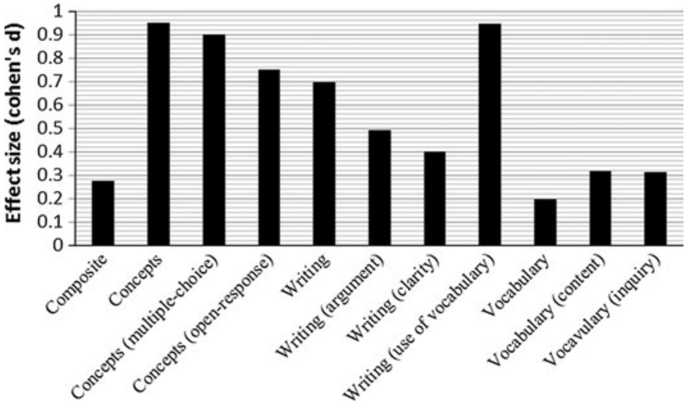

Fig. 2 Effect size comparison across achievement category

students on average gained 1.49 points (SD = 2.34) on *Science Writing*. Finally, the students' mean *Science Vocabulary* gain score was 1.32 out of a possible 30 (SD = 5.60). Effect sizes allowed us to compare gains across achievement categories. There was a high effect on *Science Concepts* (d = .949), medium effect on *Science Writing* (d = .698), and no effect on *Science Vocabulary* (d = .195). Thus, the greatest gain was on *Science Concepts.*

The sub-categories provide a more nuanced interpretation of achievement gains (see Fig. 2). Regarding *Science Concepts*, the effect size was higher for multiple-choice items (d = .900—high effect) than for constructed-response items (d = .749—medium effect). Effect size differences were also found when looking at *Science Writing*'s sub-categories: highest for *using science vocabulary* (d = .945—high effect), second highest for *argument* (d = .493—low effect), and lowest on *clarity* (d = .402—low effect). Finally, regarding *Science Vocabulary*'s sub-categories (all multiple-choice items), effect sizes were low and relatively similar between *content* and *inquiry* vocabulary (d = .316, d = .313, respectively).

Did These Gains Vary by Level of English Language Proficiency?

Table 6 shows that gain scores did vary for the four levels of English language proficiency (Beginning/Early Intermediate, Intermediate, Early Advanced/Advanced, and English Only) according to achievement category. Figures 3, 4, and 5 present comparative displays of these groups' pre-post performance by achievement category.

Descriptively, Beginning/Early Intermediate students scored the lowest on the *Science Concepts* pretest (M = 8.65, SD = 3.06), followed by Intermediate (M = 9.67, SD = 3.45), English Only (M = 11, SD = 3.17), and Early Advanced/Advanced (M = 11.33, SD = 3.17) students (see Fig. 3). By the posttest, the same relative patterns remain. However, in terms of gain scores (see Table 6), English Only students gained the least (M = 2.25, SD = 2.61), followed by Beginning/Early Intermediate students (M = 2.58, SD = 3.20) and Intermediate

Table 6 Science achievement gains by English language proficiency

Category	B/EI		I		EA/A		EO		F statistic	*p* value
	N	Gain *M* (SD)	N	Gain *M* (SD)	N	Gain *M* (SD)	N	Gain *M* (SD)		
Concepts	48	2.58 (3.20)	57	2.63 (3.22)	47	2.91 (2.53)	131	2.25 (2.61)	.634	.594
Writing	60	2.43 (2.16)	73	1.88 (2.55)	67	1.16 (2.27)	76	1.07 (2.33)	2.194	.091
Vocabulary	48	2.65 (5.03)	54	.944 (4.24)	57	1.93 (2.78)	134	2.0 (3.41)	4.059	.009**

B Beginning, *EI* Early Intermediate, *I* Intermediate, *EA* Early Advanced, *A* Advanced

* $p < .05$ ** $p < .01$ *** $p < .001$

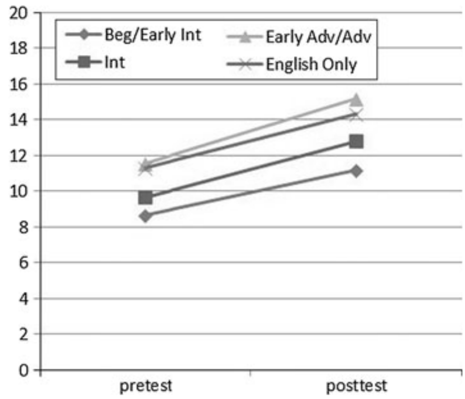

Fig. 3 Science concepts pre-posttest comparison by English proficiency

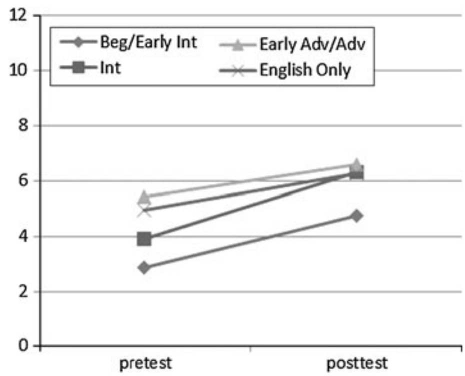

Fig. 4 Science writing pre-posttest comparison by English proficiency

Students ($M = 2.63$, SD = 3.22). Early Advanced/Advanced gained the most ($M = 2.91$, SD = 2.53).

A similar pretest pattern was found for *Science Writing*: Beginning/Early Intermediate ($M = 2.86$, SD = 1.88), followed by Intermediate ($M = 3.93$, SD = 1.87), English Only ($M = 4.95$, SD = 2.08), and then Early Advanced/Advanced ($M = 5.42$, SD = 2.02) students (see Fig. 4). The same relative pattern

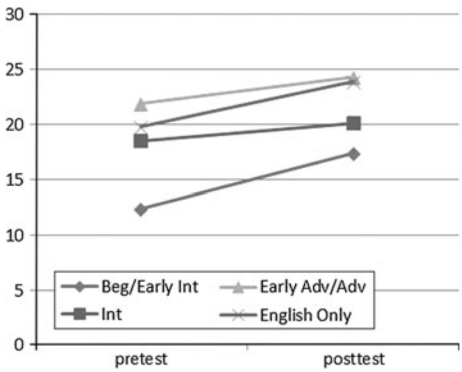

Fig. 5 Science vocabulary pre-posttest comparison by English proficiency

was found for posttest *Science Writing* scores. However, Beginning/Early Intermediate students demonstrated the highest *Science Writing* gain score ($M = 2.43$, $SD = 2.16$), followed by Intermediate ($M = 1.88$, $SD = 2.55$), Early Advanced/Advanced ($M = 1.16$, $SD = 2.27$), and English Only ($M = 1.07$, $SD = 2.33$) students.

Finally, for *Science Vocabulary* the Beginning/Early Intermediate students ($M = 12.36$, $SD = 4.78$) once again scored lower than the Intermediate ($M = 18.50$, $SD = 6.78$), English Only ($M = 19.78$, $SD = 6.75$) and Early Advanced/Advanced ($M = 21.91$, $SD = 5.26$) students on the pretest (see Fig. 5). The relative pattern of scores once again remained the same for the posttest. However, for gain scores Beginning/Early Intermediate students demonstrated the highest ($M = 2.65$, $SD = 5.03$), followed by English Only ($M = 2.00$, $SD = 3.41$) and Early Advanced/Advanced ($M = 1.93$, $SD = 2.78$) students.

Three one-way ANOVAs were calculated to test for differences in the *Science Concepts, Science Writing,* and *Science Vocabulary* gains among students of the various levels of English proficiency (see Table 6). At the $\alpha = .05$ level, the ANOVA revealed no statistical difference among the four English proficiency levels for both *Science Concepts* $F(3, 102) = .634$, $p < .594$ and *Science Writing* gain scores $F(3, 170) = 2.194$, $p < .091$. However, there was a statistical difference for *Science Vocabulary* gain scores $F(3, 118) = 4.059, p < .009$. Given this result, a tukey post-hoc test was used to determine which means among the four English proficiency groups were significantly different from each other (Field, 2013). This test revealed that students with Intermediate proficiency in English gained significantly less than those with Beginning/Early Intermediate proficiency ($p < .007$) and English Only ($p < .029$).

Discussion

Our preliminary analyses of student pre/post-test scores on the project assessment yielded a complex mix of results that will aid future analyses (both quantitative and qualitative). First, taken as a whole, all students demonstrated statistically significant learning gains whether looking at composite (i.e., all items) or achievement category (i.e., science concepts, science writing, and science vocabulary) and sub-category scores (e.g., argument, clarity, and vocabulary use). It is important to note that students of first year elementary school teachers achieved these gains.

Learning gains, however, differed across the three achievement categories. In particular, it is interesting to note the difference in the results when measuring vocabulary in terms of *definitions* (no effect size for the vocabulary achievement category) as opposed to measuring the *use* of vocabulary (large effect size for the vocabulary sub-category under science writing). This result may lend some support for the efficacy of the project's intervention, which promotes the use of vocabulary in authentic literacy tasks over decontextualized rote memorization of science vocabulary. However, item type could be an intervening variable—the vocabulary achievement category score is derived from multiple-choice items while the science writing vocabulary sub-category score is based on a student constructed response.

Results become more nuanced when disaggregating the data by level of English language proficiency. Although not statistically significant, it is interesting to note that the highest level of ELLs, namely Early Advanced/Advanced, out-performed (had higher gain scores than) English Only students in Science Concepts and Science Writing. This outcome reflects the previously discussed over-performance of re-designated ELL students (Lee et al., 2008; Shaw, 2009), and may likewise be attributed to the more focused English language and literacy support these students receive compared to the typical English Only student.

Looking across the three main achievement categories (science concepts, writing, and vocabulary) and excepting Early Advanced/Advanced students, posttest scores were lower for ELLs than English only students. However, the learning *gains* exhibited by English language proficiency groups (Beginning/Early Intermediate, Intermediate, Early Advanced/advanced) were on par with English Only students, countering researchers indicating that on a national level ELLs' achievement in science, language, and literacy lags behind that of native English speakers (Buxton, 2006; Grigg et al., 2002; Lee & Luykx, 2006; NCES, 2010).

By comparing gains within ELL groups, we found subtle differences. For example, students with Intermediate proficiency in English gained significantly less than those with Beginning/Early Intermediate proficiency and English Only students in Science Vocabulary. It may be the case that the vocabulary tested by the project assessment was already familiar to Intermediate ELL students. Also, ELLs are not a homogenous group, and teacher preparation interventions may ultimately impact students of varying levels of English proficiency in different ways.

Overall, these results offer some promise that the instruction provided by FYTs, and by extension the project's intervention, can improve ELLs' science and literacy learning, as well as learning for English only students. We should bear in mind that these gains occurred over a *single* science unit (usually 2–4 weeks of instruction), and may not capture learning that takes place over the course of an entire school year.

Limitations and Next Steps

The results should be interpreted in light of the study's inherent limitations. Some have to do with the data set itself. The 191 students reside in classrooms of nine teachers. Hierarchical Linear Modeling (HLM) analyses (see Bryk & Raudenbush, 1992) could account for this nesting; however, the relatively small number of teachers dramatically reduced the power needed to conduct HLM analysis. Also, with only one 6th grade teacher, the 191 students are not distributed equally across grade levels (although all grades are represented). While a strength of our sample is the comparable size of ELLs to non-ELLs (48–52 %, respectively), we also had to collapse two English language proficiency groups to conduct our analyses. Other limitations have to do with the state of our analyses to date. Gain scores, for example, may not be the best measure for comparison. The significance observed from the *t*-tests may not hold up with more sophisticated analyses that control for different variables. We will use tests such as analysis of covariance (ANCOVA) and ordinary least squares (OLS) linear regression to further tease out the nuances in the observed patterns. In such analyses, predictors, for example ELL status and English language proficiency, can be added to exploring models for predicting post-test scores.

In addition, we will explore relationships between fidelity of implementation of project instructional practices and student achievement. Participating teachers were observed three times while teaching the common science unit with corresponding ratings on each of the project instructional practices. We also gathered extensive qualitative data (e.g., observation field notes, audio-recordings of selected lessons, teacher debriefs, student work samples) that provide a richer context for examining and explaining the results.

Implications

This study makes an important contribution to the literature on preparing teachers to work with ELLs because it focuses on the impact of an intervention for pre-service science teachers on the learning of students in their classrooms. Most prior research examines the impact of professional development for experienced teachers on their classroom practices and ELL achievement (cf. Lara-Alecio et al., 2012). Our methodology of following pre-service teachers into the classroom as in-service teachers to gather student learning data holds high potential for informing similar studies in the future. There are two main findings reported in this paper that are important for research on improving the teaching and learning of ELLs and the preparation of the teachers who serve them.

First, this study confirms the findings of previous research that the integration of science language and literacy practices can improve the achievement of ELL in science concepts, writing and vocabulary (Bravo & García, 2004; Cervetti et al., 2007; Lee et al., 2008; Ovando & Combs, 2012; Rivet & Krajcik, 2008; Rosebery & Warren, 2008). The achievement gains of the ELLs in this study—at all levels of English Language Proficiency—were equivalent to that of native English speakers in all assessment domains. This runs counter to 30 years of research that has demonstrated that the achievement of ELLs in science language and literacy lags behind that of native speakers and that the cumulative gap increases over time (Buxton, 2006; Grigg et al., 2002; Lee & Luykx, 2006; NCES, 2010). These are important findings because they indicate that the focus on developing academic language and literacy in the subject areas emphasized in the new Common Core State Standards and Next Generation Science Standards could result in positive learning outcomes for ELLs across subject areas.

Second, the study indicates that it is possible to begin to link the practices taught in pre-service teacher preparation to novice teacher practice and student learning outcomes. Associated with this linkage are two critical components of the ESTELL pre-service teacher education program: (1) the use of explicit models of practice, and (2) the development of coherence across program components.

As noted earlier, explicit modeling of the six ESTELL teaching practices came from method instructor-led science units and mentor teachers (i.e., CTs and TSs) who themselves received training on these practices. This modeling facilitated the pre-service teachers' development of expertise through observation, analysis, and experience with the instructional approaches they were being prepared to teach (Abell & Cennamo, 2004; Goldman et al., 2007; Roth et al., 2011). Such modeling of instructional practices helps pre-service teachers identify, analyze, and subsequently use new teaching strategies by focusing their attention on specific classroom events (Abell & Cennamo, 2004; Goldman et al., 2007; Roth et al., 2011; Schwartz & Hartman, 2007; Sherin, 2004). The models also promote productive discourse for both individual and collaborative reflection (Pointer Mace, Hatch, & Iiyoshi, 2007, Sherin, 2004; Zhang, Lundeberg, Koehler, & Eberhardt, 2011), as well as help novice teachers more closely approximate the beliefs and behaviors of more experienced teachers (Ash, 2007; Segal, Demarest, & Prejean, 2006; Sherin, 2004).

Pre-service teachers also need opportunities to practice integrated instructional approaches with the student population they are being prepared to teach with intensive feedback, coaching, and support (Joyce & Showers, 1995; Loucks-Horsley et al., 1998; Speck & Knipe, 2001). Therefore, the pre-service teachers used a modified lesson plan template to design science lessons that infuse the ESTELL teaching practices for their student teaching experience, which occurred in classrooms with ELLs.

To effectively implement the ESTELL pre-service teacher education program model, it is critical to develop coherence between the instructional approach being taught in the science methods courses and the mentoring and feedback being provided to them by TSs and the CTs in whose classrooms they do their teaching practicum. The ESTELL teaching practices became a unifying framework to align method course instruction with professional development for mentor teachers—so that the pre-service teachers saw connections between coursework and their student teaching experience.

The ESTELL project's focus on language and literacy development in science instruction demonstrates the potential for significantly improving the preparation of novice teachers to teach the rapidly growing population of linguistic minority students. This is particularly important given data that demonstrate that the majority of pre-service teachers do not receive adequate preparation for teaching ELLs (Freeman et al., 2004; Menken & Antunez, 2001) and that few novice teachers feel prepared to teach ELLs (Ballantyne et al., 2008). There is an urgent need to improve this situation that results in thousands of new teachers graduating each year unprepared to teach a significant and vulnerable segment of the K-12 the student population.

The ESTELL project offers a promising model that may be broadly applicable to teacher preparation programs across the country. The requisite revamping of standard science methods courses and support for CTs do not pose insurmountable logistical or resource demands. The potential for positive impact on student learning, such as we have suggested here, is a significant source of evidence to support further dissemination of the model, thus broadening the positive impacts on the science and English language and literacy learning of ELLs.

Acknowledgments This study was conducted with support from the National Science Foundation DR K-12 program, Grant #0822402.

References

Abell, S. K., & Cennamo, K. S. (2004). Videocases in elementary science teacher preparation. In J. Brophy (Ed.), *Advances in research on teaching: Using video in teacher education* (Vol. 10, pp. 103–130). New York: Elsevier.

Aikenhead, G. (2006). *Science education for everyday life: Evidence-based practice*. New York: Teachers College Press.

Ash, D. (2007). Using video data to capture discontinuous science meaning making in non-school settings. In R. Goldman, R. Pea, B. Barron, & S. J. Derry (Eds.), *Video research in the learning sciences* (pp. 207–226). Mahwah, NJ: Erlbaum.

Ballantyne, K. G., Sanderman, A. R., & Levy, J. (2008). *Educating English language learners: Building teacher capacity* (roundtable report). Washington, DC: National Clearinghouse for English Language Acquisition. http://www.ncela.gwu.edu/publications/. Accessed 3 Oct 2013.

Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research.

Barton, A. C. (2003). *Teaching science for social justice*. New York: Teachers College Press.

Bianchini, J. A., Johnston, C. C., Oram, S. Y., & Cavazos, L. M. (2003). Learning to teach science in contemporary and equitable ways: The successes and struggles of first-year science teachers. *Science Education, 87*(3), 419–443.

Bravo, M., & García, E. E. (2004). *Learning to write like scientists: English language learners' science inquiry and writing understandings in responsive learning contexts.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Buck, G. A., & Cordes, J. G. (2005). An action research project on preparing teachers to meet the needs of underserved student populations. *Journal of Science Teacher Education, 16*(1), 43–64.

Business Higher Education Forum. (2006). *Collaborating to address the Math and science shortage: A state university—business partnership*. Washington, DC: American Council on Education. http://www.bhef.com/publications/collaborating-address-math-and-science-teacher-shortage-state-university-business. Accessed 3 Oct 2013.

Buxton, C. A. (2006). Creating contextually authentic science in "low-performing" urban elementary schools. *Journal of Research in Science Teaching, 43*(7), 695–721.

California Council on Science and Technology. (2010). *The preparation of elementary school teachers to teach science in California: Challenges and opportunities impacting teaching and learning science.* Sacramento, California: Author. http://www.ccst.us/publications/. Accessed 3 Oct 2013.

Cervetti, G. N., Pearson, P. D., Barber, J., Hiebert, E., & Bravo, M. A. (2007). Integrating literacy and science: The research we have, the research we need. In M. Pressley, A. K. Billman, K. Perry, K. Refitt, & J. Reynolds (Eds.), *Shaping literacy achievement*. New York: Guilford.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Commission on Professionals in Science and Technology. (2007). *Four decades of STEM degrees,* 1966–2004: *The devil is in the details.* STEM Workforce Data Project. CPST report no. 6. Retrieved from https://www.cpst.org/STEM/STEM6_Report.pdf.

Common Core State Standards Initiative. (2010). *Common core state standards for English language arts & Literacy in history/social studies, science, and technical subjects.* Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Darling-Hammond, L. (2006). Securing the right to learn: Policy and practice for powerful teaching and learning. *Educational Researcher, 35*(7), 13–24.

de Jong, E. J. (2004). After exit: academic achievement patterns of former English language learners. *Educational Policy Analysis Archives, 12*(50). http://epaa.asu.edu/epaa/v12n50/v12n50.pdf. Accessed 3 Oct 2013.

Doherty, R. W., & Pinal, A. (2004). Joint productive activity and cognitive reading strategy use. *TESOL Quarterly, 38*(3), 219–227.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: Sage.

Freeman, D., Garcia, E., Herrera, S., Murray, D., Valdés, G., & Walqui, A. (2004). *Preparing teachers to teach English language learners: I. What do we know*. Panel presented at the American Educational Research Association annual meeting, San Diego, CA.

Gándara, P., Maxwell-Jolly, J., & Driscoll, A. (2005). *Listening to teachers of English language learners: A survey of California teachers' challenges, experiences, and professional development needs.* UC Berkeley: University of California Linguistic Minority Research Institute. http://escholarship.org/uc/item/6430628z. Accessed 3 Oct 2013.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston, MA: Allyn and Bacon.

Goldman, R., Pea, R., Baron, B., & Derry, S. J. (2007). *Video research in the learning sciences*. Mahwah, NJ: Erlbaum.

Grigg, W. S., Daane, M. C., Jin, Y., & Campbell, J. R. (2002). *The nation's report card. Reading*, 2003-521. Retrieved from http://www.nces.ed.gov/nationsreportcard.

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

Hilberg, R. S., Tharp, R. G., & DeGeest, L. (2000). The efficacy of CREDE's standards-based instruction in American Indian mathematics classes. *Equity & Excellence in Education, 33*(2), 32–40.

Hosoume, K., & Barber, J. (1999). *Terrarium Habitats*. Berkeley: Lawrence Hall of Science, University of California.

Joyce, B., & Showers, B. (1995). *Student achievement through staff development: Fundamentals of school renewal* (2nd ed.). White Plains, NY: Longman.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lara-Alecio, R., Tong, F., Irby, B. J., Guerrero, C., Huerta, M., & Fan, Y. (2012). The effect of an instructional intervention on middle school English learners' science and English reading achievement. *Journal of Research in Science Teaching, 49*(8), 987–1011.

Lawrence Hall of Science. (2007). *Soil habitats assessment system.* Berkeley: Lawrence Hall of Science, University of California.

Lee, O., Buxton, C., Lewis, S., & LeRoy, K. (2006). Science inquiry and student diversity: Enhanced abilities and continuing difficulties after an instructional intervention. *Journal of Research in Science Teaching, 43*(7), 607–636.

Lee, O., & Luykx, A. (2006). *Science education and student diversity: Synthesis and research agenda.* Cambridge, UK: Cambridge University Press.

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.

Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners: Language demands and opportunities in relation to next generation science standards. *Educational Researcher, 42*(4), 423–433.

Loucks-Horsley, S., Hewson, P. W., Love, N., & Stiles, K. (1998). *Designing professional development for teachers of science and mathematics*. Thousand Oaks, CA: Corwin Press.

Lynch, S. (2001). "Science for all" is not equal to "one size fits all": Linguistic and cultural diversity and science education reform. *Journal of Research in Science Teaching, 38*(5), 622–627.

Lyon, E. G. (2013). Learning to assess science in linguistically diverse classrooms: Tracking growth in secondary science preservice teachers' assessment expertise. *Science Education, 97*(3), 442–467.

Menken, K., & Antunez, B. (2001). *An overview of the preparation and certification of teachers working with limited English proficient (LEP) students*. Washington, DC: National Clearinghouse for Bilingual Education.

National Academy of Sciences. (2010). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads.* A research report co-sponsored by the Committee on Underrepresented Groups and the Expansion of Science and Engineering Workforce Pipeline; Committee on Science, Engineering, and Public Policy; Policy and Global Affairs; National Academy of Sciences, National Academy of Engineering, Institute of Medicine. Washington, DC: National Academies Press. http://www.nap.edu/catalog/12984.html. Accessed 3 Oct 2013.

National Center for Education Statistics. (2010). *The condition of education 2010 (NCES 2010-028).* Washington, DC: U.S. Department of Education.

National Center for Education Statistics (2011). *The nation's report card: Science 2009.* National Center for Education Statistics. Institute of Education Sciences. Washington, DC: U.S. Department of Education.

National Clearinghouse for English Language Acquisition. (2007). *The growing numbers of limited English proficient students.* Office of English Language Acquisition, Washington DC: U.S. Department of Education. http://www.ncela.gwu.edu/search/?cx=009395468658320509053%3Ais72in2vyqy&cof=FORID%3A10&ie=UTF-8&as_q=growing+numbers. Accessed 3 Oct 2013.

National Research Council. (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas.* Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

Oakes, J., Joseph, R., & Muir, K. (2004). Access and achievement in mathematics and science: Inequalities that endure and change. In J. Banks & C. Banks (Eds.), *Handbook of research on multicultural education* (2nd ed., pp. 69–90). San Francisco: Jossey Bass.

Ovando, C. J., & Combs, M. C. (2012). *Bilingual and ESL classrooms* (5th ed.). Boston, MA: McGraw-Hill.

Parsad, B., Lewis, L., Farris, E., & Greene, B. (2001). *Teacher preparation and professional Development (NCES 2001-088).* Washington DC: U.S. Department of Education, National Center for Education Statistics,

Pointer Mace, D., Hatch, T., & Iiyoshi, T. (2007). Teaching in and teaching from the classroom: Using video and other media to represent the scholarship of teaching and learning. In R. Goldman, S. Derry, R. Pea, & B. Barron (Eds.), *Video research in the learning sciences* (pp. 397–410). Mahway, NJ: Lawrence Erlbaum Associates.

Rivet, A. E., & Krajcik, J. S. (2008). Contextualizing instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. *Journal of Research in Science Teaching, 45*(1), 79–100.

Rodriguez, A. J. (1998). The dangerous discourse of invisibility: A critique of the national research council's national science education standards. *Journal of Research in Science Teaching, 34*(1), 19–37.

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

Roth, K. J., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. (2011). Video based lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching, 48*(2), 117–148.

Santos, M., Darling-Hammond, L., & Cheuk, T. (2012). *Teacher development to support English language learners in the context of common core state standards.* Stanford, CA: Stanford University. http://ell.stanford.edu/papers/policy. Accessed 3 Oct 2013.

Schwartz, D. L., & Hartman, K. (2007). It is not television anymore: Designing digital video for learning and assessment. In R. Goldman, S. Derry, R. Pea, & B. Barron (Eds.), *Video research in the learning sciences* (pp. 335–348). Mahway, NJ: Lawrence Erlbaum Associates.

Segal, J., Demarest, E., Prejean, A. (2006). *Teaching for understanding: A guide to video resources*. Washington DC: National Education Association Research Department. http://wwwns.isea.org/assets/docs/Teaching_for_Understanding__A_Guide_ToVideo_Resources.pdf. Accessed 2 Dec 2013.

Shaw, J. M. (2009). Science performance assessment and English learners: An exploratory study. *Electronic Journal of Literacy Through Science, 8*(3). http://ejlts.ucdavis.edu/volume/8.

Sherin, M. G. (2004). New perspectives on the role of video in teacher education. In J. Brophy (Ed.), *Advances in research on teaching: Using video in teacher education, 10* (pp. 1–27). New York, NY: Elsevier JAI.

Speck, M., & Knipe, C. (2001). *Why can't we get it right? Professional development in our schools*. Thousand Oaks, CA: Corwin Press.

Stoddart, T. (2013). Restructuring pre-service teacher education to respond to increasing student diversity. *Research in Higher Education Journal, 9*, 1–14.

Stoddart, T., & Mosqueda, E. (in press). Teaching science to English language learners: A study of pre-service teacher preparation. *Teacher Education & Practice, 26*(4).

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*(8), 664–687.

Stoddart, T., Solis, J., Tolbert, S., & Bravo, M. (2010). A framework for the effective science teaching of English Language Learners in elementary schools. In D. Sunal, C. Sunal, & E. Wright (Eds.), *Teaching science with Hispanic ELLs in K-16 classrooms*. Charlotte, NC: Information Age Publishing.

Suriel, R. L., & Atwater, M. M. (2012). From the contribution to the action approach: White teachers' experiences influencing the development of multicultural science curricula. *Journal of Research in Science Teaching, 49*(10), 1271–1295.

U.S. Census Bureau. (2010). *American community survey* [2005–2009]. Retrieved from www.census.gov/acs/.

Villegas, A. M., & Lucas, T. (2002). Preparing culturally responsive teachers rethinking the curriculum. *Journal of Teacher Education, 53*(1), 20–32.

Young, J. W., Cho, Y., Ling, G., Cline, F., Steinberg, J., & Stone, E. (2008). Validity and fairness of state standards-based assessments for English language learners. *Educational Assessment, 13*(2–3), 170–192.

Zhang, M., Lundeberg, M., Koehler, M., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and Teacher Education, 27*(2), 454–462.