What Teacher Librarians Need to Know about Science Information Literacy and Second Language Learners: What Quantitative Data Doesn't Tell Us[1]

Patricia Montiel Overall, Ph.D.
University of Arizona School of Information Resources and Library Science
1515 E. First Street
Tucson, Arizona, 85718
U.S.A.
overall@u.arizona.edu

Abstract

Science is considered a critical area within the curriculum and instruction by teachers alone is not enough to ensure success for all students. School librarians must be considered in delivering science information to students, particularly those who are second language learners. This two-year study examined the effect of teacher and librarian collaboration (TLC) on inquiry-based science information literacy of Latino students in the United States. Although no significant gains were found between Control and Intervention classes, qualitative data indicated that TLC was successful in motivating students, improving inquiry and information literacy, and understanding of science concepts. The implication is that that test scores alone do not provide a complete assessment of student learning. Factors limiting student gains included reduced science and library time, and state policies that removed second language learners from regular classroom instruction. Continued advocacy for TLC is recommended to provide students needed tools for long-term academic success.

Keywords: Teacher librarians, science information literacy, second language learners

Introduction

As in many parts of the globe, the United States has increasingly become a diverse society with a growing population of underserved groups, particularly second language learners, who are falling behind their mainstream peers. In the U.S., Latinos currently comprise the fastest growing population (U.S. Department of Commerce, 2012) in which large educational gaps exist. The need to close the gap through culturally relevant classroom and library instruction is critical to support their educational and life-long success, yet they remain among the most underserved population of library users (Guëreña, & Erazo, 2000) placing them at risk and at a distinct disadvantage in an information rich and science conscious society. In addition to having overall low academic achievement (Slavin & Calderón, 2001), high drop out rates (National Assessment of Educational Progress, 2005; National Center for Education Statistics, 2011; Martínez, DeGarmo, & Eddy, 2004), and limited access to technology (Espinoza-Herold, 2003; Mossberger, Tolbert, & Gilbert, 2006). Latino students are behind in science (National Assessment of Educational Progress, 2011).

Problem Statement

Inasmuch as science is considered a critical area within the curriculum, instruction by

[1] The author would like to thank and acknowledge the Institute of Museum and Library Services for funding the research, Patricia Jones for the statistical analysis, and Anthony Hernández for advice and review.

teachers alone may not be enough to ensure success for all students. The role of school librarians must be considered in delivering science information to students, particularly those who are second language learners. Although considerable research has been undertaken to examine ways of improving science instruction for Latino students (Buxton, Lee, & Penfield, 2009/2010; Cuevas, Lee, Hart, & Deaktor, 2005; Fashola, Slavin, & Calderón, 2001; Hart & Lee, 2003; Lee, 2005; Lee & Fradd, 1998; Lee, Fradd, & Sutman, 1995; Lee, Hart, & Deaktor, 2005; Lee, Maerten-Rivera, Buxton, Penfield, Secada, 2009) limited information exists on how instruction provided by school librarians[2] working with teachers to provide jointly planned instruction in science affects Latino students' science scores. No comprehensive studies exists in science education literature on the instructional role of librarians in developing science information literate students (e.g., students who are able to effectively and efficiently find, evaluate, and use science information), although information literacy methods are considered invaluable in developing science literate students and are closely aligned with research methods used in science of formulating questions, searching for information, developing a hypothesis, collecting and analyzing data, and drawing conclusions. The purpose of this quantitative study was to examine test scores of Latino students whose teachers collaborated with librarians on developing science information literacy and inquiry-based instruction, and to explain results in light of additional factors.

Literature Review

Elementary school experiences prepare students for more complicated schoolwork in content areas particularly science. However, a considerable body of literature indicates that elementary school students lack adequate information literacy skills (Badilla Quintana, Cortada Pujol, & Riera Romaní, 2011). It is clear from student test scores (Loveless, 2013a, 2013b) and from anecdotal evidence that students also need to improve in science (ScienceInsider, n.d.). This is particularly true of Latino second language learners whose scores in science indicated a thirty two point gap between Latino and non Latino students (National Center for Education Statistics, 2011). Research indicates that several important considerations are not taken into account with second language learners: 1) the importance of first language in developing literacy, and 2) the importance of appropriate conditions (context) for second language learners to develop literacy in (August & Shanahan, & Escamilla, 2009). Wong-Fillmore & Snow (2000) suggest that educators need intensive and deep understanding of "educational linguistics" (p. 13), which would include linguistic knowledge of cognates, dialects, and phonemes.

In addition to needing to know about language acquisition, a considerable body of information exists on the effect of libraries on student achievement (Lance, 1994; Lance, Rodney, Hamilton-Pennell, 2000, 2002; Lance & Russell, 2004; Rodney, Lance, & Hamilton-Pennell, 2002). These studies indicate that when schools have adequate library resources and when librarians are available to collaborate with teachers to access these resources, student achievement on standardized tests improves. Teacher and librarian collaboration is also recommended as a means of improving student understanding of course material (Buzzeo, 2002; Donham, 2001; Haycock, 2003, 2007; Small, 2002). Within the library and information science field, collaboration between teachers and librarians is recommended as a means of connecting library and subject content instruction (American Association of School Librarians [AASL] and Association for Educational Communications and Technology [AECT], 1998). Teacher librarian collaboration (TLC) is widely promoted by prominent library and information science professionals (Callison, 1999; Doll, 2005; Eisenburg, 2008;) and is

[2] The term librarian will be used throughout the paper to avoid confusion although the term teacher-librarian is commonly used to refer to school librarians. Additional preferred terms are school library media specialist, media specialist, school resource specialist and school librarian.

arguably better than traditional library instruction as a way of improving student learning. A proposed TLC model indicates that high-end collaboration (e.g., integrating library instruction and content instruction) is most successful (Montiel-Overall, 2005; Montiel-Overall & Hernández, 2012) in improving student learning.

Method

This paper reports on quantitative data from two schools involved in a two-year mixed methods study on the effect of TLC on science information literacy of Latino students. Instructional Research Board approval and school district approval was obtained to conduct the study in two school districts in a large metropolitan area of the southwest United States with a large population of second language learners. This report presents findings from one school district in which teachers, librarians, and students participated. Students were enrolled in third, fourth, and fifth grades during the study. Teachers and librarians received extensive intervention professional development on teacher and librarian collaboration (see Montiel-Overall & Hernández, 2012 on TLC instrument), teaching science and developing information literacy skills of second language learners by collaborating on instruction of science FOSS kits (Delta Education, 2007). Quarterly benchmark tests were given to assess student progress. Standardized tests in science, reading, writing, and English were also given to students. Scores for students whose teachers did not receive professional development served as a Control. Qualitative data was also collected. Finding previously reported (Montiel-Overall & Grimes, 2013) are discussed in a later section of this paper (see Discussion below).

Student Participants

Elementary third, fourth, and fifth grade students from two schools in a rural school district participated in the study. School A was a neighborhood school in a low socioeconomic area where families who were bilingual or Spanish speaking resided.
School B was located within the same vicinity, however it was a magnet school, which drew from a wide geographic area. The majority of parents from this school were Spanish speakers. Many students spoke only Spanish at the beginning of the study.

Teacher and Librarian Participants

Four teachers (one male and three females) and two librarians (both females) from two schools participated. One female teacher and both librarians were fluent Spanish speakers.

Professional Development

Professional development workshops were held monthly during year one (Y1) and quarterly during year two (Y2). The workshops focused on four modules including science, information literacy, teacher and librarian collaboration, and language and culture. Participants were mentored by master teachers and expert librarians at four-hour workshops where teachers and librarians received instruction on collaboratively teaching inquiry-based science lessons that incorporated information literacy standards (AASL & AECT, 1998), and Standards for the 21st Century Learner (American Association of School Librarians, 2007). During the workshops and at weekly meetings between teachers, librarians, and one of the expert librarians who acted as a peer mentor to the teachers and librarians at each school, participated in jointly planned science lessons. Several models for teaching information literacy skills were presented to teachers and librarians (e.g., Big6, Information Search Process). The Big6 model (Eisenberg and Berkowitz, 1990), which identifies six steps for information problem-solving: solve information-based problems: task definition, information seeking strategies, location and access, use of information, synthesis, and evaluation, was selected by expert librarians who served as peer mentors. Expert librarians who acted as peer mentors were selected because of their extensive years of experience as librarians who collaborated with teachers.

Quarterly Science Benchmarks
Students were given quarterly science benchmark tests to evaluate students' understanding of science modules taught in 2008-2009 and in 2009-2010. Students were also given standardized tests in science, math, literacy, and English language proficiency. The tests were given to all students in the state.

Data Collection and Analysis

Analyses of data from science kits were conducted separately for 2009 and 2010 primarily because students did not remain with the same cohort (with the exception of one teacher whose class remained intact between fourth and fifth grades) and also because grade level assignments varied for teachers.

2008-2009: Table 1 summarizes statistics for the benchmark science test for each group. Scores were converted to percentages by dividing the raw score by the number of items for each test and multiplying by 100. All tests in 2008-2009 had 45 items.

Table 1. Summary Statistics for Science Benchmark Tests 2008-2009.
Mean on Top Line, Standard Deviation in Parentheses.

School	Group	Grade	N	sci01	sci02	sci03	sci04
A	**I**	**03**	22	34.04 (11.17)	38.18 (11.01)	46.77 (14.23)	47.37 (14.52)
	C	**03**	19	37.43 (14.25)	40.58 (9.71)	51.46 (11.07)	47.37 (12.06)
B	**I**	**03**	21	31.53 (8.24)	37.88 (13.80)	42.22 (13.37)	42.96 (15.58)
	C	**03**	20	32.11 (5.92)	33.33 (8.68)	40.67 12.88	36.67 16.15
A	**I**	**04**	23	40.68 (9.58)	40.97 (9.01)	48.79 (13.56)	51.98 (12.71)
	C	**04**	23	43.86 (10.12)	39.13 (7.78)	47.25 (11.97)	53.14 (9.91)
B	**I**	**04**	17	38.56 (15.71)	41.05 (12.62)	44.44 (14.25)	46.93 (13.35)
	C	**04**	20	35.00 (10.49)	33.56 (6.03)	36.55 (9.33)	39.11 (11.90)

In general, scores at School B were lower than they were at School A. Scores tended to increase through the fourth quarter, although this was not always the case.

Scores were analyzed separately by grade level using a mixed-design ANOVA with quarterly scores as a repeated measure (Q1-Q4) and Group (Intervention vs. Control), Grade (3rd, 4th), and Teacher as independent variables. Teachers were nested within Group and Grade combinations. The results of this analysis are shown in Table 2 below.

Table 2. ANOVA Summary Table for 2008-2009 Science Kit Quarterly Benchmark Scores.

Between subjects	DF	Mean Square	F Value	Pr > F
group	1	470.34	1.18	0.280
grade	1	1042.96	2.61	0.108
group*grade	1	381.60	0.95	0.330
teacher(group*grade)	4	1904.68	**4.76**	**0.001**
Error	157	399.92		

Within subjects	DF	Mean Square	F Value	Pr > F
time	3	3447.38	**64.78**	**<.0001**
time*group	3	143.02	**2.69**	**0.046**
time*grade	3	375.47	**7.06**	**0.0001**
time*group*grade	3	67.41	1.27	0.285
time*teacher(group*grade)	12	74.93	1.41	0.158
Error(time)	471	53.21		

The interactions of time with grade and time with group were significant (within subjects) as well as the effect of teacher (group*grade) (between subjects). Since the latter interaction effect is a between subjects effect, it does not examine differences among teachers at any one specific point in time. Rather, this effect examines whether average of the four test scores (by teacher) differ significantly. Pairwise t tests (using a pooled error term across all pairings) were carried out within each grade level to examine the between subjects interaction. To adjust for inflation of Type I error because of the large number of comparisons, the overall alpha level (.05) was divided by the number of tests carried out. This yielded a criterion significance level of .0041. Within the third grade, students average test score for the teacher in the control group in School B was significantly lower than the average test score for the teacher in the control group in School A (p=.0017). Since both were control group teachers, this was of interest only in that they were at different schools. Within the fourth grade, the average test score for the teacher in the control group at School B was significantly lower than the average test score for the teacher in the control group at School A (p=.0018). The teacher in the control group at School B also had significantly lower student test scores than the teacher in the intervention group at School A (p=.0015). Again, these comparisons are of less interest because they involve comparisons of teachers at different schools.

2009-2010: Scores for the 2009-2010 are shown in Table 3, below. While the design for 2008-2009 was balanced with third and fourth grade classrooms represented in both schools, the 2009-2010 data had classrooms spread across third, fourth, and fifth grades.

Fourth grade classrooms were available for the study at both School B and School A. However, third grade classrooms were available only at A and not B, and fifth grade classrooms were available only at B and not A. As a result, we carried out separate ANOVAs by grade level to examine the data. These will be described following the presentation of the summary statistics for this data set.

Scores were converted to percentages in the same way as described above.

Table 3. Summary Statistics for Science Benchmark Tests 2009-2010.
Means on First Line, Standard Deviations are Shown in Parentheses

School	**Group**	**Grade**	N	sci01	sci02	sci03	sci04
A	**I**	**03**	19	38.13 (10.69)	36.14 (12.83)	47.49 (12.13)	51.11 (15.14)
	C	**03**	19	37.54 (11.06)	42.34 (7.50)	49.24 (7.88)	53.10 (13.84)
B	**I**	**04**	17	41.31 (14.45)	43.14 (13.01)	51.76 (18.51)	49.15 (11.78)
	C	**04**	15	29.33 (9.98)	30.52 (7.34)	38.07 (10.00)	34.37 (9.27)
A	**I**	**04**	23	44.64 (13.40)	39.81 (12.33)	44.64 (14.59)	47.92 (13.81)
	C	**04**	27	42.96 (12.16)	40.33 (12.04)	46.42 (13.08)	46.75 (11.12)
B	**I**	**05**	14	41.43 (13.55)	39.73 (11.74)	46.35 (9.01)	42.22 (13.19)
	C	**05**	16	41.11 (9.39)	36.85 (10.56)	43.06 (9.32)	38.33 (9.34)

The first ANOVA examined fourth-grade students only. This allowed us to examine potential school differences without the additional problem of confounding due to inconsistent grade levels. The factors included School (A vs. B) and Group (Intervention vs. Control) as well as Time of assessment (Q1-Q4). This resulted in a mixed-design ANOVA with one within subjects factor (Time) and two between-subjects factors (School and Group). The results of the analysis are shown below in Table 4.

Table 4. Mixed Design ANOVA for Fourth Grade 2009-2010 Science Scores.
Significant Effects are Shown in Bold

Between subjects	**DF**	**Mean Square**	**F Value**	**Pr > F**
school	1	1556.21	3.29	0.074
group	1	3487.43	7.38	**0.008**
school*group	1	3347.33	7.08	**0.010**
Error	78	472.86		

Within subjects	DF	Mean Square	F Value	Pr > F
time	3	917.63	16.44	**<.0001**
time*school	3	218.25	3.91	**0.010**
time*group	3	17.00	0.30	0.822
time*school*group	3	22.21	0.40	0.755
Error(time)	234	55.83		

Four effects are significant in the analysis above. These include the between subjects effects of Group and the interaction of School and Group and the within subjects effects of Time and the interaction of Time and School.

One clear effect is that within schools, while the levels of performance differ the Intervention group at School B always scores substantially above the control group, for instance, the change in score from one test administration to the next is quite similar for the two groups. This suggests no real advantage for the intervention group at this grade level. While students in the Intervention group at B finish substantially higher than those in the Control group, they also started almost as high. For School A, the ordering of scores varied from quarter to quarter, but the differences were not great. School A students in the Intervention group scored higher than those in the Control group for the first testing, but then the Control group was slightly higher than the Intervention group at the second and third testings. Since the scores are so close, this suggests random variation from testing to testing.

There was a significant effect for Group, meaning that after the scores were averaged over schools and time periods, the overall group averages differed significantly. In addition, there was a significant effect for the interaction of School and Group. Regarding the School by Group interaction, the groups were similar for all quarterly assessment scores at School A, whereas the difference between the intervention and control group was much larger at School B. The control group at School B had much lower scores than the intervention group.

Two additional analyses were carried out to examine differences at the third and fifth grade levels. Tables 5a and 5b present the ANOVA summary tables for the analyses of the third and fifth grades at Schools A and B, respectively.

Table 5a. ANOVA Summary Table for Third Grade 2009-2010 Science Scores for Elementary School A

Source	DF	Mean Square	F Value	Pr > F
Between subjects				
group	1	207.93	0.58	0.452
Error	36	359.57		
Within subjects				
time	3	1833.92	29.88	**<.0001**
time*group	3	75.72	1.23	0.301
Error(time)	108	61.38		

Table 5b. ANOVA Summary Table for Fifth Grade 2009-2010 Science Scores for Elementary School B

Source	DF	Mean Square	F Value	Pr > F
Between subjects				
group	1	201.25	0.66	0.424
Error	28	305.99		
Within subjects				
time	3	214.74	**3.98**	**0.011**
time*group	3	18.50	0.34	0.795
Error(time)	84	53.99		

Results of both analyses are similar; the effect of Time is significant in each case, meaning that scores differed significantly between quarters. However, since groups did not differ, and the interaction of Time by Group is not significant, there was no real difference between the Intervention and Control groups either when averaged over all four time periods or when considered individually.

While the control group made steady progress throughout the year, scores for the intervention group dropped in the second quarter. After that, scores increased steadily, and were slightly lower than those for the control group. Because of the large variation in individual scores, the differences in type of progress were not reflected in a significant Time*Group interaction, but there are clearly differences in scores over time. Progress for fifth grade students was mixed. While there was variation from one period to the next, scores did not exhibit a steady upward trend for either group. Scores in the second and fourth marking periods dropped from the previous period, and neither group appeared to change more than the other; the control group started slightly lower than the intervention group and remained lower throughout the academic year.

Analysis of Total and Averaged FOSS Kit Scores

In addition to the analyses described above, the four FOSS kit scores were combined to create two different composite scores. The first was the sum of all four scores divided by 4 (i.e., a mean of four scores). These scores are referred to as "total scores" below, since they are based on a total of four scores. In this analysis, the sample size will be the same as the repeated measures analysis above, since the latter analysis includes only students who had all four scores available. The second composite score was the mean of whatever scores the student had available. The scaling for this will be consistent regardless of the number of scores included; there is one disadvantage in that when data are missing, the mean score may not represent an accurate picture of student learning. As noted above, student scores were expected to increase over the four testing periods. These scores are referred to as "average scores" below. Since the "average scores" are based on just what scores each student has available, there will be more average scores than total scores in general.

Once the composite scores were calculated, a hierarchical design ANOVA was carried out for each type of score. The effects included Group (Intervention vs. Control), Grade (3 vs. 4 for the 2008-2009 data), Group*Grade Interaction, and Teacher (Group*Grade). This latter effect was used as the error term for testing the Group, Grade, and Interaction effects (see Myers & Well, 2000) and is discussed in the analysis of the 2008-2009 data.

2008-2009: Table 6 presents the means by classroom of the total and average FOSS kit scores for the intervention and control groups.

Table 6. FOSS Composite Score Means by Group and Grade Level, 2008-2009

School	Group	Grade	N	scitot09	N	sciav09
A	**I**	**03**	22	41.59 (11.35)	25	41.16 (11.11)
	C	**03**	19	44.21 (9.56)	28	43.51 (10.05)
B	**I**	**03**	21	38.65 (11.41)	27	37.78 (11.25)
	C	**03**	20	35.69 (9.23)	28	34.93 (8.77)
A	**I**	**04**	23	45.60 (9.64)	27	45.60 (10.16)
	C	**04**	23	45.85 (8.43)	27	45.43 (8.28)
B	**I**	**04**	17	42.75 (12.65)	25	42.55 (12.82)
	C	**04**	20	36.06 (7.09)	24	36.48 (7.11)

Total scores ranged from a low of 35.69 to a high of 45.85. The mean scores ranged from 34.93 to 45.60. There was no consistency in terms of score rankings between the Intervention and Control group counterparts. The ANOVA summary tables for the analyses of these two variables are given in Tables 7a and 7b, below.

Table 7a. ANOVA Summary Table for Total FOSS Kit Score 2008-2009

Source	DF	Mean Square	F Value	Pr > F
Group	1	117.58	0.25	0.645
Grade	1	260.74	0.55	0.500
Group*Grade	1	95.40	0.20	0.678
Teacher(Group*Grade)	4	476.17	**4.76**	**0.0012**
Error	164	99.98		

Table 7b. ANOVA summary table for Mean FOSS kit score 2008-2009

Source	DF	Mean Square	F Value	Pr > F
Group	1	149.61	0.26	0.638
Grade	1	528.39	0.91	0.394
Group*Grade	1	107.41	0.19	0.6889
Teacher(Group*Grade)	4	579.22	5.70	**0.0002**
Error	203	101.61		

The results for both analyses are comparable. Only the effect of Teacher is statistically significant. As noted above, the Control group mean scores were somewhat lower than those for the Intervention students, but the differences were sometimes quite small. Post hoc tests for pairwise differences within grade and school were calculated for both types of scores (i.e., comparing Intervention and Control at each school and grade, a total of four analyses). None of the contrasts were statistically significant. The source of the significance for the effect of Teacher was due to contrasts that cross schools (e.g., third grade control teachers at each school) or within the same intervention group but at different grade levels, which are not of interest in this case.

2009-2010:

The summary statistics and ANOVA results for fourth grade students in 2009-2010 are presented below.

Table 8. Summary Statistics for Total and Averaged Science Benchmark Scores 2009-2010 for Fourth Grade Students

School	Group	Grade	N	scitot10	N	sciav10
Schools A	**I**	**04**	23	44.25 (11.58)	29	43.51 (11.98)
	C	**04**	27	44.12 (10.18)	29	44.02 (9.83)
School B	**I**	**04**	17	46.34 (13.40)	24	42.15 (13.64)
	C	**04**	15	33.07 (7.08)	25	33.22 (6.80)

While the total and averaged score means are close for both groups at School A, the between-group differences are more dramatic for School B, with a 13-point difference in total score means and a 9-point difference in averaged score means. Tables 9a and 9b present the ANOVAs for total and averaged scores, respectively.

Table 9a. ANOVA Summary Table for Composite Total FOSS Kit Score, Fourth Grade 2009-2010

Source	DF	Mean Square	F Value	Pr > F
School	1	389.05	3.29	0.0735
Group	1	871.86	7.38	**0.0081**
School*Group	1	836.83	7.08	**0.0095**
Error	78	118.22		

Table 9b. ANOVA Summary Table for Composite Average FOSS Kit Score, Fourth Grade 2009-2010

Source	DF	Mean Square	F Value	Pr > F
School	1	982.90	8.36	**0.0047**
Group	1	469.76	4.00	**0.0483**
School*Group	1	590.81	5.02	**0.0271**
Error	103	117.57		

The School*Group interaction term was statistically significant for both the total and averaged score means. Post hoc tests were carried out using the Bonferroni adjustment ($\alpha\beta$=.0083). Control and Intervention did not differ significantly for School A for the total scores (p=.965), but the difference was statistically significant for School B (p=.0009). While the two Control groups differed significantly (p=.0023), the difference between the two Intervention groups was not significant (p=.550). For the averaged scores, the results were the same for the comparison of Control and Intervention within schools; the difference at School A was not significant, but the difference at School B was (p=.0048). In addition, while the Control groups differed significantly between schools (p=.0004), the Intervention groups did not (p=.457).

Relationship between English Proficiency and Academic Performance

In the final sets of analyses, language proficiency scores were used as classification variables and classroom performance in the science benchmark tests and the SSMS classroom scores were examined to determine the extent to which English proficiency influenced academic outcomes.

2008-2009: Students measured on the language proficiency exam fell into one of four categories: Emergent (1 student or 1%), Basic (6 students, 6%), Intermediate (43 students, 46%), or Proficient (44 students, 47%). Students at the Emergent or Basic levels were grouped into a single category representing Low proficiency, and students who were untested were assumed to be native English-speaking students. Scores on achievement measures such as the science benchmarks, the averaged scores, and the SSMS measures were then analyzed in one-way ANOVAs to determine whether mean scores differed among the English-proficiency groups. Results are presented in Table 10. Because so few students fell into the two lowest categories, their results were dropped from the analyses, and results are shown only for Intermediate, Proficient, and native-English speakers. A posteriori analyses were carried out to determine which groups differed significantly when an ANOVA was significant. The Bonferroni procedure was used to adjust the criterion significance level by the number of contrasts (.05/3=.0167). Superscripts next to mean scores indicate homogeneous groups. For example in the row labeled "Benchmark 01", the score for the Intermediate group has a superscript a, whereas the Proficient and native English-speaking groups are labeled with a b. This indicates that the mean for the first science benchmark for the Intermediate group is significantly lower than both the score for the Proficient and native-English speaking groups, which do not differ significantly from each other. Direction of differences is obtained by inspection of the mean scores.

Table 10. Analysis Results for Each Performance Measure by English language Proficiency, 2008-2009.
Columns 2-4 Present Means, Standard Deviations, and Sample Sizes; Column 5 Presents F Statistic and Significance Level for One-way ANOVA

Measure	Intermediate	Proficient	Native	F, p(F)
Benchmark 01	28.8 [a] (8.1) 39	36.7 [b] (10.9) 40	40.0 [b] (11.2) 112	**16.28** **<.0001**
Benchmark 02	31.0 [a] (7.6) 39	37.9 [b] (10.3) 39	41.5 [b] (9.7) 105	**17.56** **<.0001**
Benchmark 03	35.7 [a] (10.2) 40	46.4 [b] (12.1) 43	48.6 [b] (12.2) 101	**17.49** **<.0001**
Benchmark 04	34.3 [a] (11.4) 40	46.8 [b] (13.7) 42	49.8 [b] (12.3) 100	**22.28** **<.0001**
Total science	32.5 [a] (7.2) 35	42.0 [b] (9.9) 36	45.2 [b] (9.3) 91	**24.97** **<.0001**
Average science	32.4 [a] (6.8) 43	42.3 [b] (9.9) 44	44.1 [b] (10.0) 117	**24.28** **<.0001**
SSMS Reading SS	413.7 [a] (36.2) 41	462.7 [b] (36.6) 43	478.7 [b] (47.9) 105	**33.51** **<.0001**
SSMS Math SS	427.3 [a] (41.6) 41	469.3 [b] (37.6) 43	482.2 [b] (50.1) 105	**21.26** **<.0001**
SSMS Writing SS	436.6 [a] (42.6) 41	472.7 [b] (35.9) 43	488.2 [b] (39.3) 104	**25.29** **<.0001**
SSMS Science SS (4 only)	452.7 [a] (49.3) 15	488.1 [b] (29.9) 17	521.6 [c] (41.7) 64	**18.72** **<.0001**

All analyses were significant (p<.0001) and with one exception, the Intermediate proficiency students were significantly different from both the Proficient and native English-speaking students. The sole exception was on the SSMS Science test, where the difference between the Intermediate and Proficient group was marginally significant (p=.0174) and the difference between the Proficient and native English-speaking students was highly significant (p=.0036). This is almost certainly an artifact of the very small sample sizes for the science scores, since the differences between means are all at least 30 points. Since the Bonferroni adjustment tends to be conservative and lose power, all three groups should be considered as differing significantly here.

2009-2010: Table 11 gives mean scores for each measure broken down by English language proficiency. No students were rated as less than Intermediate on the language proficiency exam and students without language proficiency scores were assumed to be native English speakers.

Table 11. Analysis Results for Each Performance Measure by English Language Proficiency, 2009-2010. Columns 2-4 Present Means, Standard Deviations, and Sample Sizes; Column 5 Presents F Statistic and Significance Level for One-way ANOVA

Measure	Intermediate	Proficient	Native	F, p(F)
Benchmark 01	28.2 [a] (7.0) 24	38.4 [b] (11.0) 40	41.7 [b] (12.1) 119	**14.27** **<.0001**
Benchmark 02	30.7 [a] (6.2) 25	38.9 [b] (9.6) 45	39.2 [b] (12.2) 122	**6.45** **.0019**
Benchmark 03	34.4 [a] (12.4) 20	46.3 [b] (10.3) 42	47.8 [b] (13.0) 113	**9.97** **<.0001**
Benchmark 04	33.9 [a] (8.7) 26	45.1 [b] (14.0) 45	47.5 [b] (14.2) 116	**10.75** **<.0001**
Total science	31.8 [a] (5.0) 19	42.7 [b] (8.8) 37	45.1 [b] (10.7) 93	**14.65** **<.0001**
Average science	31.8 [a] (5.5) 27	42.1 [b] (9.2) 45	43.1 [b] (11.8) 93	**12.70** **<.0001**
SSMS Reading SS	427.9 [a] (28.7) 26	471.6 [b] (29.1) 45	482.5 [b] (29.1) 120	**19.47** **<.0001**
SSMS Math SS	333.2 [a] (22.0) 26	380.6 [b] (32.3) 45	384.2 [b] (38.2) 120	**14.65** **<.0001**
SSMS Writing SS (5 only)	426.5 (51.1) 6	446.5 (40.7) 8	485.4 (70.3) 33	2.83 .0696
SSMS Science SS (4 only)	455.9 [a] (32.6) 13	515.6 [b] (35.8) 22	511.0 [b] (42.1) 62	**11.44** **<.0001**

All analyses were statistically significant except for SSMS Writing. Typically, the Intermediate group scored significantly lower than the Proficient and the native English-speaking groups. The scores for the group with the Proficient rating were somewhat lower than the native English-speaking group, although not substantially so. The SSMS writing test was administered only to students in grade 5. Although there are large between-group differences, the analysis was not significant. There are two reasons for this. The first is that the sample sizes in the non-native speaking groups are small (6 and 8, respectively). The second is that the standard errors are quite large for the Intermediate and Native groups (51 and 70, respectively). These two factors result in a loss of power and a lack of statistical significance, even in the face of large group differences. For the science SSMS, the Proficient group had a slightly higher mean than the native-speakers. This suggests that

English proficiency, at least at this level, was not a factor in academic achievement in science. For the most part, the results are quite similar between the two years and it is clear that English proficiency has a substantial effect on academic performance.

Scores for the FOSS average and total (combined) scores were also correlated with SSMS and language proficiency outcomes. Table 12 shows the results for 2008-2009 and 2009-2010.

Table 12. Correlations of Total FOSS Scores with SSMS and Language Proficiency Outcomes for 2008-2009 and 2009-2010

	Total FOSS	**Total FOSS**	**Total FOSS**
2008-2009	Grade 3	Grade 4	
Reading	**0.83 <.0001 82**	**0.76 <.0001 83**	
Math	**0.69 <.0001 82**	**0.71 <.0001 83**	
Writing	**0.54 <.0001 82**	**0.41 0.0001 83**	
Science	**	**0.83 <.0001 83**	
2009-2010	Grade 3	Grade 4	Grade 5
Reading	**0.75 <.0001 37**	**0.88 <.0001 82**	**0.75 <.0001 30**
Math	**0.83 <.0001 37**	**0.72 <.0001 82**	**0.72 <.0001 30**
Writing	**	**	**0.73 <.0001 30**
Science	**	**0.83 <.0001 81**	**

All correlations were statistically significant regardless of year, content area, or grade level. Correlations for average and total FOSS benchmarks (refer to discussion above for explanation of difference in computational method to obtain scores) were quite similar, as would be expected. Correlations between the Writing test and the FOSS benchmarks were the lowest of the group in either year, ranging from a low of .41 (Grade 4, 2008-2009) to a high of .73 for Grade 5 in 2009-2010. However, this latter value is based on 30 observations, and the correlation for the counterpart average score was .10 lower. Correlations with the state's standards measure of science (SSMS) measure were typically high; only one correlation was below .80. It should be noted, however, that the correlation of the FOSS benchmarks composite scores with Reading and Math were similarly high, so it is likely that they reflect general ability as much as knowledge of science.

Discussion

Based on two years' results from students' FOSS kit science benchmark scores, the intervention did not appear to have had a measurable impact on student achievement. Fourth grade scores indicate that both the Control and Intervention classes improved over time with no statistical difference between the groups, although fourth grade Intervention students at School B obtained higher benchmark scores than Control students in Y1 for all testings, whereas test scores tended to be both closer and reverse in ranking across quarters at School A. At both grade levels the scores for School A were typically higher than those for School B. (see Table 2). Although scores for Control and Intervention at both schools increased over time, they fell below 2011 national scores in science (Loveless, 2013a, 2013b). Fourth grade performance for both Control and Intervention started at less than 40% accuracy and were well below 60% accuracy on final tests. Fifth grade scores showed mixed results and no classroom mean exceeded 50% indicating that the test are very difficult for students.

The lack of significant gains for the Intervention group could be the result of several factors. First, students who were second language learners were not performing at the same level as non-second language learners. Although efforts were made to ensure equal numbers of English speakers and non-English speakers in the Intervention and Control classes, most students in the Intervention classes were non-English speakers. A further discussion of this is discussed below (see *Other Factors*).

Second, the intervention was indirect inasmuch as professional development was provided to teachers and not students. In addition, teachers and librarians may not have focused on developing information tested. Additional direct observation of teachers and librarians would have confirmed that what was presented to teachers and librarians in the professional development workshops was taught to students. Third, two years may not be long enough to see significant gains in students' scores. It may take longer to move students to higher levels of achievement and to change the culture of the classroom from direct instruction to inquiry-based instruction (Supovitz & Turner, 2000).

Finally, although the benchmark tests were designed to evaluate students' understanding of the science through activities presented in the FOSS kits, test questions may not have captured what students were being taught or what they were learning. For example, although science activities included exercises to develop students' inquiry skills, the benchmarks evaluated content.

Other Factors

To fully understand findings from this report, qualitative data provides information about other factors that help explain the lack of significant gains by students of Intervention teachers and librarians. A state mandate implemented during Y2 required second language learners who had not passed a language proficiency test to be assigned daily to four-hour blocks of instruction in English grammar, vocabulary, language arts, and writing in a classroom other than that of the Intervention teacher. Thus, time spent on science by Intervention teachers was severely reduced (e.g., in the remaining two hours, science, math, social studies, geography, and history were taught). Also during Y2 due to budgetary cuts, library instruction was reduced from one hour three times a week to once a week for half an hour.

In spite of non-significant results from quantitative data, information obtained from qualitative data previously reported (Montiel-Overall & Grimes, 2013) reveals that the Intervention had extremely positive effects on student interest, motivation, and understanding of science, which are not evident in results of quantitative data. Teachers and librarians focused on standards for second language learners and incorporated cognates, visuals aids, children's literature, graphic organizers, vocabulary walls, weather maps, and science websites, to engage second language learners in science. Librarians in particular used students' first language, which researchers have found to improve science for English language learners (Buxton, Lee, & Penfield, 2009/2010). The following comments by participating teachers and librarian illustrate the success of TLC on science teaching and learning for themselves as well as students, which was not evident in test scores:

"The kids love it [science]. I mean I have one group doing an experiment…other group is sitting down with me on reading…like sequencing" [P11:320]

"[Students] aren't just sitting there waiting for somebody to give them some information, they're actually trying to figure it out…you can see that light in their eyes just come alive. It's like, 'Oh, my god! I've learned this one.' " [P10:105]

"[The librarian] finished up a lesson on "How to Use the Library." She took half of the class to explain what is available in the library, how to log on to the computer, how to use website to help them find what they need in regards to finding books that they will need for our science lessons. I worked out an outline on how we were going to break up the [science] kit. The librarian will work with half of my class. They will use strategies taught to locate books of interest regarding Sounds." [Journal]

Implications

Several implications for library and information science (LIS) professionals result from this study. First, the use of quantitative data only to determine student learning must be used with caution. While test scores provide a means of tracking scores in achievement, over time, it is equally important to track and understand factors affecting teaching and learning to provide a complete picture of second language learners' ability. Qualitative data is an important means of providing a comprehensive analysis (Montgomery, 2011) of students' learning and interest to provide a more robust assessment of what students know and understand. Qualitative data also provides information about successful teaching strategies and methods used to achieve this goal of higher test scores.

In addition, this research suggests that the results of intervention may not be apparent within two years. Effects of intervention may be delayed and not become apparent until much later. Finally, the study highlights the effect of language on test results. Research on language and thought indicates that a strong relationship exists between language, experiences, and cognition (Boroditsky, 2011). Becoming language proficient and adjusting to a different "cognitive toolkit" (Boroditsky, 2011, p. 64) requires more time for second language learners and may require the use of multiple means to evaluate students' academic achievement in science and in other subjects.

Conclusion

This study illustrates that multiple factors contribute to student achievement, and that quantitative data alone provides an incomplete assessment of learners' ability. Factors such as neighborhood, language ability, time spent in science and in the library as well as political mandates affect student progress and success. This study illustrates that despite challenging conditions, second language learners showed steady improvement throughout a two-year study with few exceptions, and that school librarians in partnership with teachers contributed to the success.

School librarians are increasingly called upon to work with second language learners, and to provide effective and culturally relevant instruction. Integrating library instruction (e.g., information literacy, learning to do research, and finding resources) with science content is an opportunity to reinforce conceptual as well as linguistic development of second language learners. Inasmuch as time was a significant factor in student gains, school librarians must advocate for increased library time for students, particularly for second language learners who often have limited experience with libraries. Librarians are central in providing additional support for second language learners and continued advocacy for greater collaboration with teachers should continue so that TLC become becomes an integral part of all students' education.

References

American Association of School Librarians. (2007). Standards for the 21st century learner. Retrieved from http://www.ala.org/ala/mgrps/divs/aasl/guidelinesandstandards/learningstandards/AASL_Learning_Standards_2007.pdf

American Association of School Librarians & Association for Educational Communications and Technology. (1998). *Information power: Building partnerships for learning.* Chicago, IL: American Library Association.

August, D., Shanahan, T., & Escamilla, K. (2009). English language learners:
Developing literacy in second-language learners—Report of the national literacy panel on language-minority children and youth. *Journal of Literacy Research, 41*(4), 432-452. doi: 10.1080/10862960903340165.

Badilla Quintana, M. G., Cortada Pujol, M. Riera Ronaní, J. (2011). Internet navigation and information search strategies: How do [sic]children are influenced by their participation in an intensive ICT project. *International Journal of Technology and Design Education, 22*(4), 513-529. doi: 10.1007/s10798-011-9158-4.

Boroditsky, L. (2011). How language shapes thought. *Scientific American, 304*(2), 62-65.

Buxton, C., Lee, O., Penfield, R. (2009/2010). Developing English literacy through science instruction. *The International Journal of Foreign Language Teaching,* Winter, 11-13.

Buzzeo, T. (2002). *Collaborating to meet standards: Teacher/library media specialist partnerships for K–6.* Worthington, Ohio: Linworth.

Callison, D. (1999). Keywords in instruction: Collaboration. School Library Media Activities Monthly, *15*(7), 37−39 (ERIC Document Reproduction Service No. EDEJ608483).

Cuevas, P., Lee, O., Hart, J., & Deaktor, R. (2005). Improving science inquiry with elementary students of diverse backgrounds. *Journal of Research in Science Teaching, 42*(3), 337-357.

Delta Education. (2007). FOSS (Full Option Science System): Grades K-5. Nashue, NH: School Speciality, Inc. Retrieved from http://www.delta-education.com/science/foss/index.shtml

Doll, C. (2005). *Collaboration and the school library media specialist.* Lanham, MD: Scarecrow Press.

Donham, J. (2001). The importance of a model. In J. Donham, K. Bishop, C. C.

Eisenberg, M. (2008). Information literacy: Essential skills for the information age. *Journal of Library and Information Technology, 28*(2), 39-47.

Eisenberg, M., & Berkowitz, R. (1990*). Information problem solving: The Big Six skills approach to library & information skills instruction.* Norwood, NJ: Ablex.

Espinoza-Herold, M. (2003). *Issues in Latino education: Race, school culture, and the politics of academic success.* Boston, MA: Allyn and Bacon.

Fashola, Olatokunbo S., Slavin, R. E., & Calderón, M. (2001). Effective programs for Latino students in elementary and middle schools. In R. Slavin & M. Calderón (Eds.). Effective programs for Latino students. (pp. 1-66). Mahawah, NJ: Lawrence Erlbaum Associates, Inc.

Fradd, S. H., Lee, O., Sutman, F. X. & Saxton, M. K. (2002). Promoting science literacy with English language learners through instructional materials development: A case study. *Bilingual Research Journal, 25(*4), 417-439.

Guëreña, S., & Erazo, E. (2000). Latinos and librarianship. *Library Trends, 49*(1), 138-181.

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.

Haycock, K. (2003). Collaboration: Because student achievement is the bottom line. *Knowledge Quest, 32*(1), 54.

Haycock, K. (2007). Collaboration: Critical success factors for student learning. *School Libraries Worldwide, 13*(1), 25-35.

Kuhlthau, & D. Oberg (Eds.), *Inquiry-based learning: Lessons from Library Power* (pp. 13–30). Worthington, OH: Linworth.

Lance, K. C. (1994). The impact of school library media centers on academic achievement. *School Library Media Quarterly, 22,* 167-172.

Lance, K. C., Rodney, M., & Hamilton-Pennell, C. (2000). *How school librarians help kids achieve standards: The second Colorado study.* San Jose, California: HI Willow Research and Publishing. (ED 445 698)

Lance, K. C., Rodney, M., & Hamilton-Pennell, C. (2002). *How school librarians help kids achieve standards: The second Colorado study.* San Jose, California: HI Willow Research and Publishing. (ED 445 698)

Lance, K.C., & Russell, B. (2004). Scientifically based research on school libraries and academic achievement: What is it? How much do we have? How can we do it better? *Knowledge Quest, 32*(5), 13-17.

Lee, O. (2005). Science education with English language learners: Synthesis and research agenda. *Review of Educational Research, 75*(4), 491-530.

Lee, O., & Fradd, S. H. (1998). Science for all, including students from non-English language backgrounds. *Educational Researcher, 27*(4), 12–21.

Lee, O., Fradd, S. H., & Sutman, F. X. (1995). Science knowledge and cognitive strategy use among culturally and linguistically diverse students. *Journal of Research in Science Teaching, 32*, 797-817.

Lee, O., Maerten-Rivera, J., Buxton, C., Penfield, R. D., & Secada, Walter G. (2009). Urban elementary teachers' perspectives on teaching science to English language learners. *Journal of Science Teacher Education*, *20*(3), 263-286. doi: 10.1007/s10972-009-9133-z

Loveless, T. (2013a). The 2013 Brown Center Report on American Education: How Well are American students learning? Washington, DC: The Brown Center on Education Policy. Retrieved from http://www.brookings.edu/~/media/Research/Files/Reports/2013/03/18%20brown%20center%20loveless/2013%20brown%20center%20report%20web.pdf

Loveless, T. (2013b). The latest TIMSS and PIRLS scores. Part I of the 2013 Brown Center Report on American Education: How Well are American students learning? Retrieved from http://www.brookings.edu/research/reports/2013/03/18-timss-pirls-scores-loveless

Martínez, C.R., DeGarmo, D.S., & Eddy, J.M. (2004). Promoting academic success among Latino students. *Hispanic Journal of Behavioral Sciences, 26*(2), 128-151.

Montgomery, S. E. (2011). Quantitative vs. qualitative – do different research methods give us consistent information about our users and their library space needs? *Library and Information Research, 35*(111), 73-86.

Montiel-Overall, P. (2005). Toward a theory of collaboration for teachers and librarians. *School Library Media Research, 8*. Retrieved from http://www.ala.org/aasl/aaslpubsandjournals/slmrb/slmrcontents/volume82005/theory

Montiel-Overall, P. & Grimes, K. (2013). Teachers and librarians collaborating on inquiry-based science Instruction: A longitudinal study. *Library and Information Science Research, 35*(1), 41-53.

Montiel-Overall, P. & Hernandez, A. (2012). The Effect of Professional Development on Teacher and librarian collaboration: Preliminary findings using a Revised instrument, TLC-III. *School Library Research, 15*. Retrieved from http://www.ala.org/aasl/sites/ala.org.aasl/files/content/aaslpubsandjournals/slr/vol15/SLR_EffectofPDonCollaboration_V15.pdf

Mossberger, K., Tolbert , Caroline J., & Gilbert, M. (2006). Race, place, and information technology. *Urban Affairs Review, 41*(5), 583-620. doi: 10.1177/1078087405283511.

Myers, Jerome L. & Wells, Arnold D. (2003). *Research design and statistical analysis.* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum Associates.

National Assessment of Educational Progress. (2005). The nation's report card. Science. Retrieved from http://nationsreportcard.gov/science_2005/s0111.asp

National Center for Education Statistics. (2011). The nation's report card: Science 2009 (NCES 2011-451). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2009/2011451.pdf

Rodney, M. J., Lance, K. C., & Hamilton-Pennell, C. (2002). *Make the connection: Quality school library media programs impact academic achievement in Iowa.* Bettendork, IA: Mississippi Bend Area Education Agency. Retrieved March 17, 2011, from http://www.iowaaeaonline.org/vnews/display.v/ART/492b02e0d63b8

ScienceInsider. (n.d.). What's to blame for poor science test scores? *American Scientist.* Retrieved from http://www.americanscientist.org/science/pub/-1644

Slavin, R. E., & Calderón, M. (Eds.). (2001). Effective programs for Latino students. Mahwah, NJ: Lawrence Erlbaum Assoc.

Small, R. (2002). Collaboration: Where does it begin? *Teacher Librarian, 29*(5), 8-11.

Supovitz, Jonathan. A. & Turner, Herbert M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching, 37*(9), 963-980.

U.S. Department of Commerce. (2012). U.S. Census Bureau projections show a slower growing, older, more diverse nation a half century from now. Census Bureau Newsroom, 2010. Retrieved from http://www.census.gov/newsroom/releases/archives/population/cb12-243.html

Wong Fillmore, L. & Snow, C.E. (2000). What teachers need to know about language. ERIC ED-99-CO-0008

Biographical note

Patricia Montiel Overall is an Associate Professor at the University of Arizona. Her research interests include teacher and librarian collaboration and equity of access issues specifically related to Latino students.