Project MSSELL
Middle School Science for English Language Learners

Executive Summary
November 2010

Rafael Lara-Alecio, Ph. D., Principal Investigator
Texas A&M University

Beverly J. Irby, Ed. D., Principal Investigator
Sam Houston State University

Fuhui Tong, Ph. D., Internal Evaluator
Texas A&M University

Rosalinda Rodriguez, Ed. D., Aldine Area Superintendent
Aldine Independent School District

Cindy Guerrero, M. Ed., Project Coordinator
Aldine Independent School District
cguerrer@aldine.k12.tx.us

Executive Summary
Project MSSELL
First-Year Implementation and Results

The purpose of this study was to evaluate the effectiveness of an instructional intervention in academic science among 5th grade English language learners (ELLs) after one year of placement with comparison to the typical science instruction in the district where the larger research project is being implemented. This part of the study is inclusive of the first year of implementation of a two-year longitudinal experimental study.

Significance of the Study

Even though the best program for ELLs at this point remains undetermined due to lack of randomized trial studies, many states are blindly adopt approaches for ELLs, while these students "frequently confront the demands of academic learning through a yet-unmastered language" (Lee, 2005, p. 492). This study is significant in that most studies have not actually observed bilingual/ESL and English immersion science classrooms in a large scale study to take into account instructional factors in the learning of English (Bruenig, 1998; Meyer, 2000; Irby, Tong, Lara-Alecio, & Rodriguez, in press). Furthermore, to-date, only a few experimental studies have intervened to enhance classroom instruction in science for ELLs (August, Branum-Martin, Cardenas-Hagan, & Francis, 2009; Lee, Maerten-Rivera, Penfield, LeRoy, & Secada, 2008; Lynch, Kuipers, Pyke, & Syesze, 2005). The study presented here will help address the lack of research on the impact of the science classroom in relation to acquiring academic English science language among ELLs by investigating the effects of an NSF-funded enhanced instructional science model on the science achievement of Grade 5 ELLs as measured by district benchmark assessments in science.

Methods

Context and Participants

Our study was derived from a longitudinal, field-based, randomized research project targeting native both ELLs and non-ELLs in an urban school district in Southeast Texas. Over 45% of students in the district are served whose first language is Spanish. The majority of students in the selected school district site qualify for free or reduced-lunch. This particular district was selected for study because of its (a) positive reputation based on student achievement and national awards such as the Broad Foundation Prize, (b) lengthy experience working with ELLs, (c) consistency in program philosophy and implementation, and (d) because of the ease of access to English learning and regular programs within the same school throughout the district. For the purpose of this study, only ELLs were included for analysis. All participating ELLs were identified as limited English proficient with Spanish as the primary language spoken at home.

Design

Texas state law (Texas Education Code, 1995) prohibits random selection and assignment on the basis of individual students; therefore, in the larger research project, four intermediate schools from the selected district site were randomly assigned to conditions, resulting in two treatment (enhanced practice) and two control (typical practice) schools. Both ELLs and non-ELLs in the same school receive the same practice to allay contamination between experimental and control classrooms. Hence, this overall project was a quasi-experimental design at the student level and an experimental design at the school level. When a school was assigned, teachers with ELLs or non-ELLs from that campus were then randomly selected to the assigned condition within that campus. In this present study, there are 4 teachers and 166 students with consent in the treatment condition; and 8 teachers and 81 students with consent in the control condition. The distribution of all students across conditions is similar to that of students with consent only. Therefore, internal validity is secured. The post-test is about to be administered.

Intervention

The intervention is composed of two main components: (a) Teacher professional development (professional portfolio assessment, bi-weekly staff development sessions, and monthly staff meetings for paraprofessionals); and (b) Student instructional intervention implemented in 8 classrooms in 2 schools, inclusive of (a) academic science intervention (5th grade)--5E model for instruction (85 minutes daily of science instruction), (b) reading expository science text for English literacy and language acquisition (focus on vocabulary development and extension), (c) use of individual science notebooks to help students process science content through use of written academic science vocabulary. (Students are asked to predict, record, organize, draw, question, and reflect in the notebooks.), (d) technology tools (document cameras and projector, interactive white-boards, educational software, including EduSmart Science, and internet resources), (e) scientist mentors and Science Saturdays, and (f) family take-home activities and one 45-minute training on how to implement these types of science activities at home. Finally, there is a daily tutoring by trained paraprofessionals for lowest achieving students.

Measure

The benchmark science test given in the district (where the larger research project is implemented) was used to compare students' performance. The science benchmark test is a criterion-referenced test which employs a cut-off score to determine if the student passes or meets commended performance of the test. Such scoring procedure is reflective of the Texas Assessment of Knowledge and Skills (TAKS) (end-of-year Texas state standardized test) because both the benchmark tests and TAKS are aligned with the state content standards in science. On the TAKS, scaled score of 2099 is still considered failing with only 1 point below 2100 (the cut-off score) while 2101 is considered passing with only 1 point above. Therefore, to most accurately present the preliminary findings, we report the passing rate for each of the tests given every 6 weeks. In addition, we will also report students' performance on science TAKS. The TAKS, a criterion-referenced assessment, measures student mastery of the content areas of state curriculum outlined in the Texas Essential Knowledge and Skills (TEKS). The TAKS science assessments are first administered during Grade 5. As is described by TEA (2006), students who pass TAKS science in 5th grade demonstrate satisfactory performance with an at or above state passing standard and a sufficient understanding of the TEKS-aligned science curriculum. The level of commended performance reveals high academic achievement, considerably above state passing standard; and a thorough understanding of the TEKS science curriculum.

Data Sources and Analysis

Data were collected in the fall and spring of school year 2009-2010. Science benchmark test was administered to each student every 6 weeks during the school year with a total of 6 tests. Each 6-week benchmark test has a different topic area, including Physics (test 1), Chemistry (test 2), Mid-term (cumulative of physics, chemistry, and space, test 3), Earth/Space (test 4), Life Science (test 5), and final test (cumulative of physics, chemistry, space and life science, test 6). Because of our access to the district data, we also included scores from students without consent so as to determine if there is a selection bias. Results presented in this study were from students with consent only because selection bias was not detected.

Chi-square test of independence was conducted to compare the rate of passing and commended performance between intervention and control groups of ELLs. Note that it may not be beneficial to plot progress based on the scaled score, because each six weeks test cover completely different science content, which suggests that a higher score does not necessarily indicate progression, instead, it may mean that the science content is simply different, more interesting to the students, or more difficult by topic area in a single six-weeks.

Results from Benchmark Tests in Science

Chi-square test of independence was conducted to compare the passing rate between experimental and control groups for ELLs and for non-ELLs respectively. The effect size in the form of Cramer's V together with the descriptive statistics of percentage of passing and commended performance by language status and condition is presented in Table 1.

The Aldine Independent School District's science benchmark test is a criterion-referenced test which employs a cut-off score to determine if the student passes or meets commended performance of the test. Such scoring procedure is reflective of the TAKS (end-of-year Texas state standardized test) because both the benchmark tests and TAKS are aligned with the state content standards in science. On the TAKS, scaled score of 2099 is still considered failing with only 1 point below 2100 (the cut-off score) while 2101 is considered passing with only 1 point above. Therefore, to most accurately present the preliminary findings of Project MSSELL, we report the passing rate for each of the five tests (benchmark test 5 was optional due to the time conflict with TAKS administration). Results suggest that there is a statistically higher percentage of passing (4 out of 5 BM tests) and commended performance (4 out of 5 BM tests) in experimental groups than in control groups for both ELLs and non-ELLs. Such a difference is more evident among non-ELLs. For example, there is an average of 95% passing rate in the experimental group. Note that to avoid selection bias, as well as the access to the district database, scores from students with and without consent were analyzed. A visual representation of the difference between experimental and control groups are illustrated in Figures 1 and 2.

Table 1.

Preliminary Chi Square Findings from District Benchmark Science Tests by Language Status and Condition

Passing (%)						
	MSSELL ELLs			MSSELL Non-ELLs Low SES		
Test	Experimental	Control	Effect size[a]	Experimental	Control	Effect size[a]
1	85.7	84.3	0.019	95.3	91.3	0.07
2	88.4	78.9	.128*	97.8	83.6	.201**
3	83.3	79.7	0.045	93.5	73.2	.229**
4	89.4	76.6	.173**	94.7	71.8	.260**
6	85.5	65.3	.235**	97.6	73.2	.279**
Commended Performance (%)						
	MSSELL ELLs			MSSELL Non-ELLs Low SES		
Test	Experimental	Control	Effect size[a]	Experimental	Control	Effect size[a]
1	40	33.1	0.069	44.2	41.5	0.024
2	51.6	34.1	.169*	74.7	47.8	.250**
3	32.6	23.4	.097*	30.8	33.3	-0.025
4	47	34.6	.121*	66	36.1	.278**
6	41.6	25.8	.159*	56	32	.221**

*$p < .05$, **$p < .001$.
[a]positive effect size indicates higher performance in experimental condition, and a Cramer's V larger than .2 suggests a moderate degree of magnitude of difference.

Fig. 1. Comparison of % of passing and commended performance from District Benchmark Science Tests (BM3)

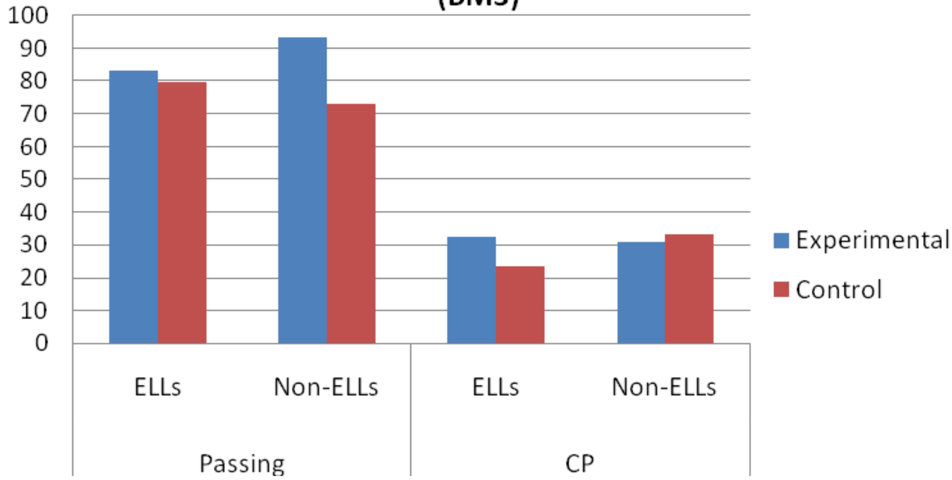
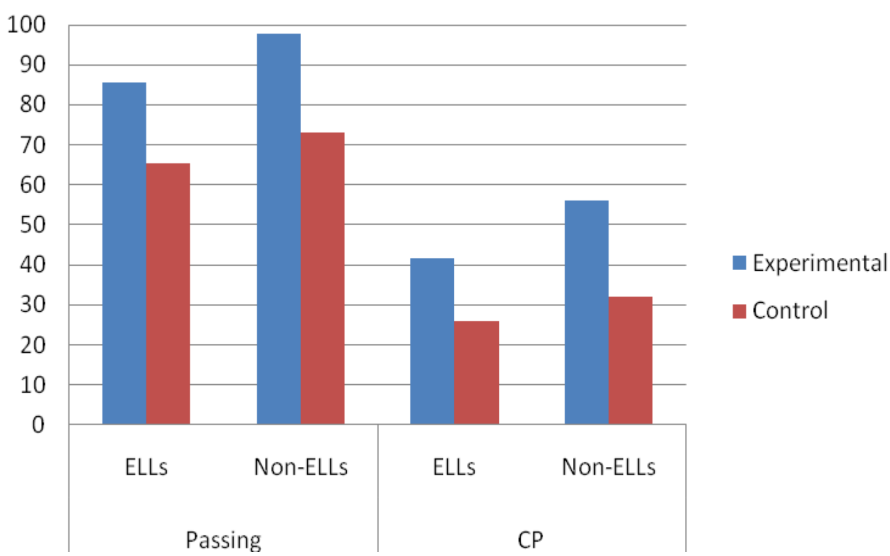


Fig. 2. Comparison of % of passing and commended performance from District Benchmark Science Tests (BM6)



Naglieri Nonverbal Ability Test (NNAT)

Nonverbal ability test was given to all MSSELL students with and without consent (n = 601). No statistically significant difference was observed between experimental and control groups for both ELLs and non-ELLs.

	ELLs			Non-ELL Low SES		
Condition	Mean	Std. Deviation	N	Mean	Std. Deviation	N
Control	97.28	11.154	139	98.12	12.852	188
Experimental	97.05	12.674	187	95.20	10.658	87
Total	97.15	12.032	326	97.19	12.257	275

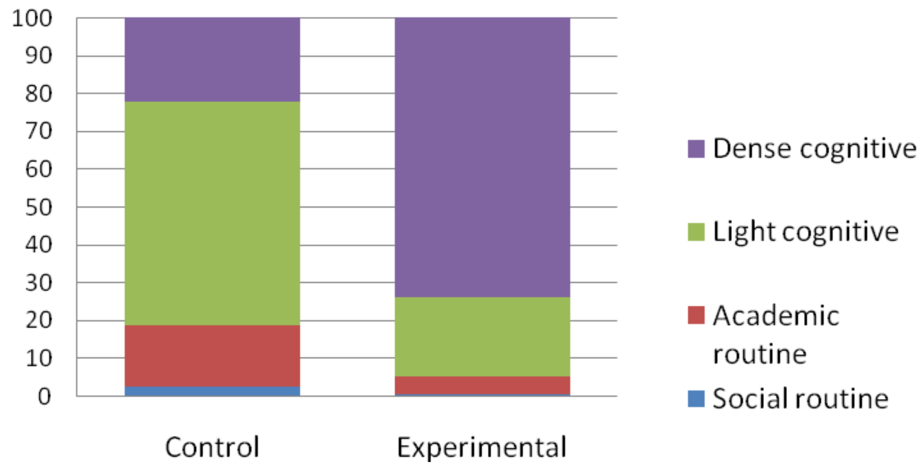
Transitional Bilingual Observation Protocol (TBOP)

There are a total of 1966 rounds of observation conducted in both experimental and control classrooms during science instruction. The data collected are frequency data in nature, and therefore, a chi-square test of homogeneity of proportion was employed to identify if the proportion of each category under every domain of TBOP is homogenous across condition. For example, in the domain of *language of instruction*, we performed the chi-square test to identify if the frequency of occurrence of each category (e.g., L1, L2) was homogenous across control and experimental classrooms. In the case when the null hypothesis of homogeneity or equal proportion ($H0$: $p1 = p2 = p3 = p4$) was rejected, a post hoc pairwise comparison was performed when necessary by examining the difference between two chi-square values calculated based on the cell values of the contingency table statistics. Unlike multiple post hoc t test procedures, which inflate α level (Type I error), a chi-square test of homogeneity maintains α at a constant level throughout the significance tests (Cox & Key, 1993). Cramer's V was also reported as type of effect size in our study (Rea & Parker, 1992).

Results are presented by specific domains in TBOP:

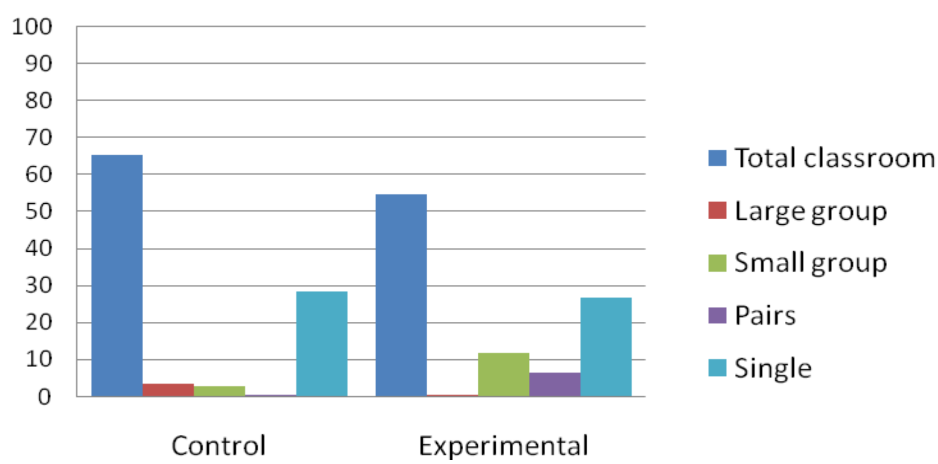
Language content. The chi-square test was significant ($p < 0.001$), with a Cramer's V of 0.48, indicating that there was statistically significant difference among teachers' time allocation in the language content between the two conditions, and the magnitude of such difference was strong. Figure 7 demonstrates a higher percentage of social language, academic routines and light cognitive observed in control classrooms than in experimental classrooms, while a higher percentage of dense cognitive content was observed in experimental classrooms. Post hoc pairwise comparisons between experimental and control classrooms within each of the four levels yielded statistically significant differences (ps < 0.01)

Fig. 7. Language content by condition



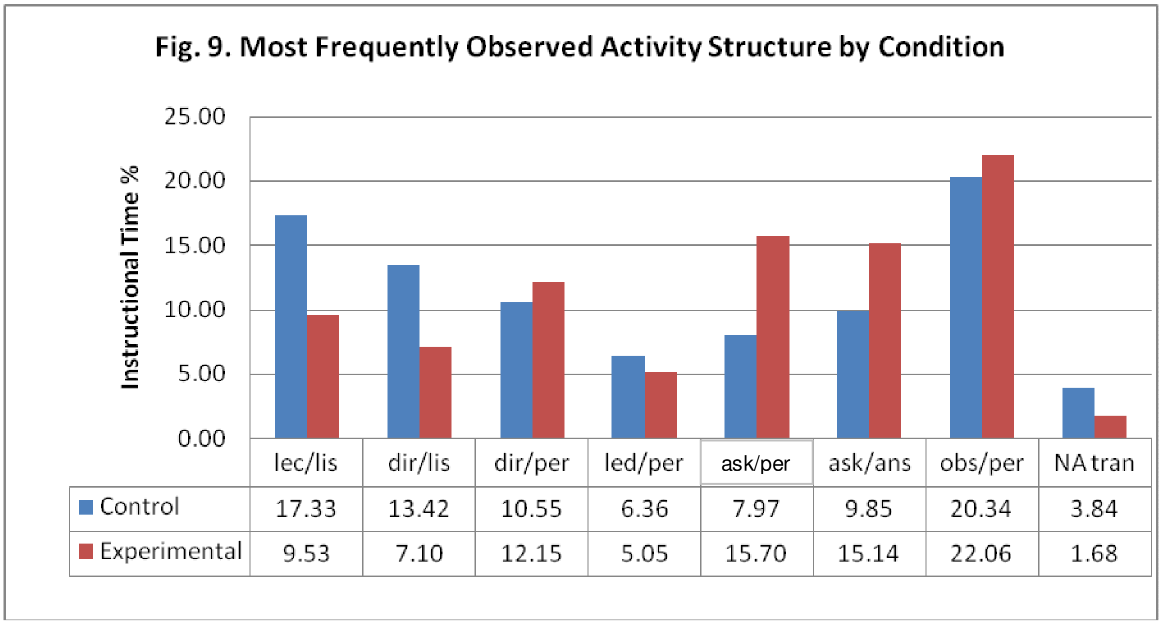
Physical group. The chi-square test was significant ($p < 0.001$), with a Cramer's V of 0.28, indicating that there was statistically significant difference among teachers' time allocation in the physical group between the two conditions, and the magnitude of such difference was moderate in strength. Figure 8 demonstrates a higher percentage of small group and pairs activities observed in experimental classrooms than in control classrooms, while a higher percentage of total classroom and large group instruction observed in control classrooms.

Fig. 8. Physical Group



Language of instruction. No statistically significant difference was observed in teachers' time allocation in the language of instruction because all teachers were speaking English during 100% of the instructional time.

Activity structure. The chi-square test was significant ($p < 0.001$), with a Cramer's V of 0.29, indicating that there was statistically significant difference among teachers' time allocation in the activity structure between the two conditions, and the magnitude of such difference was moderate in strength. Figure 9 presents the most frequently observed activity structures in both groups. Teachers in control classrooms seemed to spend more time in lecturing and directing with students listening; while teachers in experimental classrooms spent more time leading with students performing and answering questions asked by their teacher.



	lec/lis	dir/lis	dir/per	led/per	ask/per	ask/ans	obs/per	NA tran
Control	17.33	13.42	10.55	6.36	7.97	9.85	20.34	3.84
Experimental	9.53	7.10	12.15	5.05	15.70	15.14	22.06	1.68

References

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

Bruenig, N. A. (1998). Measuring the instructional use of Spanish and English in elementary transitional bilingual classrooms. *Dissertation Abstracts International*, 59(04), 1046A.

Lee, O. (2005). Science education with English language learners: Synthesis and research agenda. *Review of Educational Research*, *75*(4), 491-521.

Lee, O., Maerten-Rivera, J., Penfield, R., 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*, 31-52.

Lynch, S., Kuipers, J., Pyke, C., & Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching, 42*, 921-946.

Meyer, L. (2000). Barriers to meaningful instruction for English learners. *Theory into Practice*, *39*(4), 228-236.

Texas Education Agency. (2006). TAKS performance level descriptors. Austin, TX: Author. Retrieved from http://www.tea.state.tx.us/index3.aspx?id=3222&menu_id=793