tesol QUARTERLY
tesol international association

Developing Academic Language in English Language Learners Through Sheltered Instruction

DEBORAH J. SHORT, CAROLYN G. FIDELMAN, AND MOHAMMED LOUGUIT
Center for Applied Linguistics
Washington, DC, United States

This article describes a study examining the effects of Sheltered Instruction Observation Protocol (SIOP) model instruction on the academic language performance of middle and high school English language learners. The SIOP model is an approach for teaching content curriculum to students learning through a new language. Teachers employ techniques that make the content concepts accessible and also develop students' skills in the new language. Using a quasi-experimental design, the research was conducted in content area and English as a second language classes in two districts in northern New Jersey over 2 years. The analysis presents student achievement data from state-mandated language proficiency tests in the final year of the intervention, after most of the treatment teachers had completed their professional development in the SIOP model. There were statistically significant differences in the average mean scores in favor of the treatment student group on Writing, Oral Language, and Total English scores of the IDEA Language Proficiency Tests with small to medium effect sizes. The results from this study show that the SIOP model offers a promising approach to professional development that can improve the quality of instruction to English language learners and increase their English language achievement.
doi: 10.1002/tesq.20

Methods for second language education across the globe have adapted over time to deepen and broaden a learner's ability to use a new language. The deepening has sought to advance proficiency, and the broadening has added academic and other goals to social language abilities. Many changes have been derived from learner needs and interests as well as educators' explorations of better ways to teach. In the United States, however, another source for change is present. The educational reform movement, a political construct rather than a learner need, has had a direct impact on English language learner

(ELL) education. States now implement high-stakes testing and standards-based instruction for all students, regardless of their proficiency in English. Classroom instruction is guided by mandated standards of proficiency for core subjects such as mathematics, science, and social studies, but in many content classes little or no accommodation is made for the specific language development needs of ELLs. This lack of accommodation raises a significant barrier to success, because ELLs are expected to achieve high academic standards in English and, in many states, must pass end-of-course tests or high school exit exams in order to graduate.

Unfortunately, ELLs in elementary and secondary schools have experienced persistent underachievement on the high-stakes accountability measures enacted as a result of the No Child Left Behind (NCLB) legislation passed by Congress. This legislation requires states to assess ELLs in English language proficiency annually; in mathematics and reading each year in Grades 3-8 and once in high school; and in science once in elementary, middle, and high school. As a result, a seventh-grade beginning-level ELL who has been in the United States for 1 year would take the same mathematics test in English as a seventh-grade native English speaker who has been in U.S. schools all his or her life. On nearly every measure of state and national assessments, ELLs lag behind their native-English-speaking peers and demonstrate significant achievement gaps (Kindler, 2002; Kober et al., 2006; National Center for Education Statistics, 2009a, 2009b; Reardon & Galindo, 2009). These results are not unexpected, because most ELLs must take subject area tests using English before they are proficient in the new language.

Because these testing practices are unlikely to end soon, interventions for addressing the ELL performance gap are needed. One strategy is targeted teacher development, so that the resulting instruction provides ELLs with access to the core curriculum and concurrently develops their academic English proficiency in relation to that curriculum. The practice of integrating language development with techniques to make curricular topics more comprehensible to ELLs is generally known as *sheltered instruction* in the United States; this is when the main focus is on a specific subject curriculum and the instructor is a content specialist. It is referred to as *content-based English as a second language (ESL)* when the main focus is on language learning and the instructor is a language specialist (Crandall, 1993; Short, 2006; Snow, Met, & Genesee, 1989). Content-based language instruction is applied internationally in foreign and second language settings (including postsecondary institutions) for language minority and majority students. It is similar to content and language integrated learning (CLIL), sustained content language teaching, English for specific purposes,

and other labels (Lyster, 2007; Mohan, Leung, & Davison, 2001; Murphy & Stoller, 2001; Stoller, 2004). In the United States, it involves second language learners (usually minority language speakers) who are studying content in the new language (usually the majority language), which is the medium of instruction.

As methods for integrating language and content instruction have developed over the past two decades, many combinations of techniques have been applied to the delivery of sheltered instruction and content-based language instruction. The study presented in this article investigated one particular approach of sheltered instruction, known as the Sheltered Instruction Observation Protocol (SIOP) model. The SIOP model offers a system for lesson planning and delivery that incorporates best practices for teaching academic English and provides teachers with a coherent method for improving the student achievement. Teachers integrate instruction of content concepts with academic language in order to develop student skills in reading, writing, listening, and speaking. The concepts and language skills are aligned to state proficiency standards, and teachers use techniques designed to make academic topics accessible to students and to enable them to practice the use of academic language as it is employed in each subject area.

SIOP MODEL DEVELOPMENT AND RESEARCH

SIOP began as an observation tool for researchers to measure teachers' implementation of sheltered instruction techniques. It evolved into a lesson planning and delivery approach, known as the SIOP model, through a 7-year research study sponsored by the Center for Research on Education, Diversity and Excellence (CREDE) and funded by the U.S. Department of Education. It began in 1996 and involved teams of researchers and middle school teachers in three school districts. At that time, there was no model of sheltered instruction. Instead, teachers selected from a wide range of techniques and activities to design lessons as they chose. As a result, sheltered instruction varied widely across schools and within different classrooms in the same school.

Our goal was to create and test a model that teachers could implement with consistency and with the knowledge that doing so would improve student performance. So the teams identified best practices for integrating language and content from the professional literature and organized combinations of these techniques to build a model of sheltered instruction. During the first 4 years, the researchers analyzed teacher implementation and student effects as the teachers field-tested

variations of the model in their classrooms. In 2000, the SIOP model format was finalized with 30 features of instruction grouped into eight components: Lesson Preparation, Building Background, Comprehensible Input, Strategies, Interaction, Practice and Application, Lesson Delivery, and Review and Assessment (Echevarria, Vogt, & Short, 2000). See Figure 1 for an example of the features of the Strategies component.

The SIOP model became a framework for lesson planning and teaching. If teachers implement it with fidelity, research-supported features of instruction for ELLs will be present in their lessons. It is not, however, a step-by-step approach. Instead, it accommodates variation in teaching style. Teachers may accomplish their language and content learning goals in ways suited to the particular lesson, asking students to, for example, hypothesize the results of a science experiment with peers using cause–effect sentences, argue a position on the use of taxation to reduce a deficit, or write a journal entry about a text-to-self connection after reading a short story. The SIOP model incorporates features recommended for high-quality instruction for all students, such as cooperative learning and reading comprehension strategies (Genesee, Lindholm-Leary, Saunders, & Christian, 2006) and adds specific features for ELLs, such as the inclusion of language objectives in each lesson, oral language practice, and the development of background knowledge and academic vocabulary.

The researchers refined the SIOP model observation protocol with a 5-point scale for each of the 30 features so that the level of implementation could be measured for any lesson, ranging from 4 (*recommended practice*) to 0 (*no evidence of the practice*). Researchers and coaches could rate teachers' lessons over time using this protocol and provide explicit feedback to help them implement the model more closely. A study was conducted to establish the validity and reliability of the protocol. Four experienced professors of sheltered instruction from three universities (who were not part of the research) used the protocol to observe and score videos of teachers engaged in sheltered instruction. A statistical analysis revealed an interrater agreement of .90, and additional analyses indicated the SIOP instrument is a highly reliable and valid measure of sheltered instruction (Guarino et al., 2001).

Strategies

13. Ample opportunities provided for students to use **learning strategies**
14. **Scaffolding techniques** consistently used, assisting and supporting student understanding (e.g., think-alouds)
15. A variety of **questions or tasks that promote higher order thinking skills** (e.g., literal, analytical, and interpretive questions)

FIGURE 1. SIOP strategies component features.

The researchers examined the model's effects on student performance in the school districts using a quasi-experimental design. Because the CREDE study preceded the NCLB legislation, most of the ELLs in the research site districts were exempted from state testing. As a result, the writing assessment instrument used in the Illinois Measurement of Annual Growth in English (IMAGE) was employed as the outcome measure of academic literacy. The IMAGE was the standardized test of reading and writing that Illinois used to measure the annual growth of these skills for ELLs in Grades 3 and higher. It was found to be valid and reliable and had correlational and predictive value for achievement scores on the IGAP (Illinois' former assessment for reading and mathematics; Illinois State Board of Education, 2004). It provided subscores for five dimensions of writing—language production, focus, support/elaboration, organization, and mechanics—and a total score for each student.

Researchers administered the IMAGE writing exam in the 1998–1999 school year with a pretest in the fall and a posttest in the spring. Two groups of ELLs in Grades 6–8 sheltered classes participated: students whose teachers were trained in the SIOP model (the treatment group) and a similar group of ELLs in the same district programs whose teachers had no exposure to the SIOP model (the comparison group). Both groups included students with a range of English proficiency levels and a number of different first languages.

Results showed that ELLs in sheltered classes with teachers who had been trained in the SIOP model improved their writing and outperformed the students in comparison classes (Echevarria, Short, & Powers, 2006). The results revealed statistically significant mean differences in favor of students with SIOP-trained teachers for the total writing score and for three subtests (language production, organization, and mechanics). The treatment group also made gains over the comparison group in the focus and support/elaboration subtests, but the gains did not reach statistical significance. These results indicated that the SIOP model offered a promising approach for helping ELLs develop academic literacy skills needed for success in school.

THE NEW JERSEY SIOP STUDY

The CREDE research and development project created and field-tested the SIOP model. The study reported here was the next step in the research process; it explored the effects of SIOP model implementation on the academic literacy performance of secondary school ELLs. It extended the investigation to a new state and to school districts not part of the original research and broadened the scope to include high

school. It also examined the SIOP professional development program for teachers, which lasted 2 years (from the 2004–2005 school year through 2005–2006). Because it was conducted after NCLB was established policy, we used the state's standardized English language proficiency test as the outcome measure. The study was funded by the Carnegie Corporation of New York and the Rockefeller Foundation.

Research Question

The main research question investigated in this article is: Will ELLs in one district with teachers who received professional development in the SIOP model show significantly higher achievement in reading, writing, and oral proficiency in English on a standardized measure than ELLs in a comparable district with teachers who had no SIOP professional development? The study also sought to answer a second question: Do teachers reach high levels of implementation of the SIOP model during a sustained professional development program after 1 year or after 2 years?

Research Design

Following the guidelines set forth by the National Academy of Sciences (Shavelson & Towne, 2001), the study employed a quasi-experimental design across two school districts (one treatment, one comparison) in northern New Jersey by comparing groups of teachers and students in middle and high school and analyzing student English language achievement data. Conducting the research in one school district was not possible due to high school schedules that did not allow students to be separated into two distinct groups as well as differences between the middle schools, with one having many more ELLs and a lower overall socioeconomic status. The districts were matched as closely as possible on five contextual and participant factors: student diversity in linguistic and cultural backgrounds, size of student population, socioeconomic status, student achievement scores on state exams, and language program design at the middle and high school levels. We recruited a representative sample of teachers in both districts, whose students, in turn, became the unit of analysis for the study.

The outcome measure was the IDEA Language Proficiency Tests (IPT), a standardized assessment of English language proficiency that New Jersey used for NCLB accountability purposes. Although we report IPT scores from the baseline year and Year 1 of the intervention in our descriptive analysis, our inferential results are focused on

Year 2, when SIOP professional development was completed. At that time most of the teachers were adequately trained in the model, and we posited that by then the SIOP treatment might have a meaningful impact on student achievement.

Participants

School districts. The treatment and comparison districts were located in northern New Jersey near a large city and were representative of a growing number of metropolitan school districts that are smaller, urban-like sites of secondary migration and some primary immigration for ELL families. Each district has one large high school (Grades 9–12) and two middle schools (Grades 6–8); the three in each district participated. The treatment district served approximately 10,000 students; the comparison district roughly 6,000. The districts had diverse racial, ethnic, and linguistic student populations. Both districts had well-established ESL programs in the secondary schools and bilingual programs in some elementary schools. In both districts, ELLs experienced a persistent performance gap on state tests when compared to native English speakers.

Both locations had considerably higher levels of poverty than the state average. At the time of the study, New Jersey categorized districts into eight District Factor Groups (a measure of poverty) to determine allocation of state education funds. The comparison district was ranked at the second lowest socioeconomic level, the treatment district at the third lowest. One middle school in the treatment district was a Title I school, meaning it received additional funding due to a high number of low-income students. No comparison school had Title I status.

Teachers. Middle and high school teachers participated in both districts. Most of the teachers were voluntary participants, although some new hires in the treatment district in Year 2 were assigned to the project. In the treatment district, two cohorts of teachers were involved. Cohort 1 began in the first year of the intervention (2004–2005), and 35 teachers from this cohort remained in the study for 2 years. Cohort 2 joined in the second year (2005–2006), and 23 teachers participated. The treatment teachers taught mathematics, science, social studies, language arts, ESL, special education, and technology.

The comparison district did not have cohort groups because no SIOP professional development was involved. However, there was some teacher turnover. Twenty-three teachers participated in the first year and 22 in the second, but only 19 teachers participated across both

years. The comparison teachers taught mathematics, science, social studies, and ESL.

In both districts, approximately half of the teachers taught at the high school level and the other half at middle school. The majority of the teachers were female and White (see Table 1). The teachers had a wide range of experience, although in both districts more than 50% were veteran teachers with 10 or more years of experience. Staff turnover was low in both districts (see Table 2).

Regarding working with ELLs, Treatment Cohort 1 had, on average, 7.3 years' experience teaching these learners, and Cohort 2 had 8.1 years. Yet the experience varied widely within Cohort 1. Nineteen

TABLE 1
Demographic Data of Treatment and Comparison Teachers

	Treatment cohort 1	Treatment cohort 2	Comparison
Total no. of teachers at end of study	35	23	19
School level			
High	17	13	9
Middle	18	10	10
Subject area			
Mathematics	9	11	5
Science	5	5	6
Language arts	3	3	0
Social studies	7	3	4
ESL	9		4
Technology	2		
Special education		1	
Gender			
Male	7	6	5
Female	28	17	14
Race			
White	33	21	17
Asian	2	2	0
Black		1	2
Hispanic ethnicity			
Hispanic	5	1	1

TABLE 2
Percentage of Teachers by Years of Teaching Experience in Treatment and Comparison Districts

Teaching experience	Treatment cohort 1	Treatment cohort 2	Comparison
1 year experience or less	10%	12%	6%
2–3 years	15%	12%	0%
4–9 years	23%	12%	18%
10–19 years	31%	44%	53%
20 years or more	21%	20%	24%

Cohort 1 teachers had no previous training to work with ELLs, and eight had some training through workshops or meetings. In contrast, six Cohort 1 teachers had or were pursuing ESL teaching certification, and six had master's degrees in ESL, TESOL, or applied linguistics. This advanced training can be explained by the fact that all the ESL teachers who participated in the study were in Cohort 1. For Cohort 2, two teachers had ESL endorsements (additional coursework in ESL beyond content teaching certification), three had received some training through state or district workshops, and one had training for working with refugees. None had a master's degree in ESL, TESOL, or applied linguistics.

The comparison teachers had, on average, 7.7 years' experience working with ELLs; however, most did not provide specific information about their professional backgrounds. The four ESL teachers in the comparison district had ESL certificates or endorsements.

Students. Students who were in the ESL programs in Grades 6–12 in both districts were included in our data collection. We collected baseline data (i.e., scores on the IPT test) on the ELLs in the district schools in Spring 2004, before the study began. At that time, 386 ELLs were in the ESL program in Grades 6–12 in the treatment district, and 176 ELLs were in the comparison district's program. In 2004–2005 (Year 1 of the intervention), we included 278 ELLs in the treatment district, that is, those in the ESL program who had at least one SIOP-trained teacher. The comparison district had 169 ELLs in participating teachers' classes. In 2005–2006 (Year 2), there were 267 ELLs in the treatment district with at least one SIOP-trained teacher and 168 ELLs in the comparison district with at least one participating teacher. Over the course of the three test administrations, we collected data from 931 IPT tests (described in the Measures section) in the treatment district and 513 tests in the comparison district.

The treatment and comparison districts designated the students by the ESL proficiency levels of the IPT test. A student might be a Level 1 student, a Level 2 student, and so on for programming purposes. (Level 6 students, the most advanced, have exited—that is, transitioned out of—the ESL program.) In treatment schools, 5%–8% of the total student population was identified as limited English proficient; in comparison schools, it was 5%–7%. Table 3 displays the percentage of treatment and comparison students by ESL proficiency level based on IPT scores of the prior school year.

As seen in Table 3, the distribution of students across the levels was fairly similar for the two intervention years in both districts. However, in 2005–2006 in the treatment district, a greater percentage of students reached exited status (from 25% to 38%) and a smaller

TABLE 3
Percentage of Treatment and Comparison Students Classified by ESL Proficiency Level

	Treatment			Comparison		
	03–04	04–05	05–06	03–04	04–05	05–06
Level 1	10%	2%	2%	2%	5%	2%
Level 2	17%	8%	6%	13%	10%	15%
Level 3	30%	21%	13%	26%	18%	19%
Level 4	39%	25%	23%	34%	21%	29%
Level 5	5%	16%	14%	25%	23%	25%
Exited		−25%	38%		−21%	8%
Missing		−3%	4%		−5%	2%

percentage was at the advanced Levels 4 and 5 (from 41% to 37%). The opposite occurred in the comparison district. The percentage of exited students diminished (from 21% to 8%), but the percentage of students at advanced Levels 4 and 5 increased (from 46% to 54%).

Both districts had multilingual populations. In the treatment district, 15 native languages were represented in 2004–2005 and 18 languages in 2005–2006. The students reported 35 countries of origin. In the three treatment schools, native Spanish speakers were 24%–41% of all the secondary students, Polish speakers were approximately 6%, and Arabic speakers were 5%–7%. Other languages included Gujarati, Turkish, Pilipino, and Albanian. The number of students who lived in homes where English was not the first language ranged from 43% to 49% across the three schools. Many of these students were former ELLs and were still acquiring academic English literacy and struggled in their content classes (J. Kusielewicz, personal communication).

The comparison district students represented eight native language backgrounds and 25 countries of origin. In the three comparison schools, native Spanish speakers were 2%–17% of all the secondary students, Polish speakers were 1%–8%, and Portuguese speakers were 1%–3%. Other languages included Arabic, Gujarati, Haitian Creole, and Slovak. Like the treatment schools, the comparison schools also had a large number of former ELLs who had exited the ESL program but were still acquiring academic English.

In the treatment district, the gender breakdown was approximately 51% male and 49% female. In the comparison district it was 44% male and 56% female. More than 50% of the students in the treatment district reported Hispanic ethnicity, compared to 42% of the students in the comparison district.

Although there were some differences in the population and obtained samples of teachers and students between and within the treatment and comparison districts, we were satisfied that the two

districts were adequately similar to each other for the purpose of this quasi-experimental study.

Measures

The outcome measure was the New Jersey assessment of English language achievement, the IPT, which provided separate scores for Writing, Reading, and Oral Language. The tests were used in each of the six study schools as part of the state testing for NCLB accountability purposes.[1] The IPT has been designed to evaluate proficiency in English for nonnative speakers from 3 years old through Grade 12 (Ballard & Tighe, 2011). The three tests—Writing, Reading, and Oral Language—may be combined into a Total English proficiency score to provide an overall student assessment of English language ability, as was done by these districts. The tests were normed on a group of limited-English-proficient students representing a broad range of ethnic and socioeconomic backgrounds and language abilities.

The Writing tests have three sections: Conventions (multiple-choice sentence completion for items such as capitalization and punctuation), Write a Story (based on a picture), and Write Your Own Story (based on a prompt). Reading tests include multiple-choice items in several areas: Vocabulary (words to describe a picture), Vocabulary in Context (words to complete a sentence), Reading and Understanding (comprehension questions about a passage), Reading for Life Skills (comprehension questions about labels and popular print), and Language Usage (questions about grammar and punctuation). In these districts, ELLs took the Language Usage portion of the Reading test instead of the Conventions portion of the Writing test. In this study, two levels of the Writing and Reading tests were used: IPT 2 (Grades 4–6) and IPT 3 (Grades 7–12). Writing samples were scored with rubrics and exemplars of student writing for each level.

The Oral Language test assesses syntax, morphological structure, lexical items, phonological structure, comprehension, and oral production. The test is individually administered, and examinees progress through levels of difficulty until they reach their proficiency ceiling. Two levels of the Oral tests were used: IPT 1 (Grades K–6), and IPT 2 (Grades 7–12).

[1] In Year 2 of the study, New Jersey switched language proficiency tests from IPT to ACCESS for ELLs (WIDA Consortium, 2006). However, the ACCESS test could not provide results in a timely manner, so both districts also administered the IPT in Spring 2006. Thus we collected data from three administrations of the IPT, in 2004 (baseline year), 2005 (Year 1), and 2006 (Year 2).

The Reading and Writing tests report raw scores that can be converted to proficiency level scores (1–6). The Oral Language test reports only proficiency level scores. New Jersey has districts report proficiency level scores and sets the cutoff points. In our analyses we used the proficiency level scores. These proficiency levels relate the scores across the different language domains and grade-level tests so that the scores are valid across the grade levels. The districts use the proficiency level scores to determine student promotion within and to exit from the ESL program.

SIOP INTERVENTION

Professional Development

The SIOP professional development intervention in the treatment district had several dimensions: (a) summer institutes and workshops for coaches and teachers conducted by the researchers, (b) school-based coaching and class observations by researchers and site coaches, and (c) technical assistance via electronic media. The professional development program focused on instructional strategies for teaching academic English literacy skills and content knowledge to ELLs.

During workshops, researchers presented the SIOP model to teachers and coaches and discussed collaboratively with them effective ways to bring the model into classroom instruction, using a participatory approach with hands-on activities to demonstrate techniques that integrate language and content, cooperative mini-projects, and integration of research findings and theoretical principles. Activities were drawn from the SIOP model book (Echevarria, Vogt, & Short, 2004) and a SIOP model professional development manual (Short, Hudec, & Echevarria, 2002). In addition, to help participants learn the model and observe SIOP instruction, videos that illustrate SIOP model implementation in classrooms with a variety of grade levels, English proficiency levels, and subjects (Hudec & Short, 2002a, 2002b) were viewed and analyzed. Using sample lessons and templates, researchers and coaches also helped teachers write SIOP lesson plans. (For more training information, see www.cal.org/siop.)

Each cohort of treatment teachers had 7 days of professional development in its first year of participation to learn and apply the SIOP model: 3 days in August before school began and 4 during the year. Cohort 1 had 3 additional days in Year 2. Three part-time coaches supported the teachers in Year 1, one per school, with approximately one period daily assigned to coaching activities. Two coaches had had SIOP staff development prior to the study. The third learned along

with the teachers. All were experienced ESL educators. Two more coaches were added at the high school in Year 2. They were ESL teachers from Cohort 1. To assist with implementation, coaches and researchers observed teachers and gave feedback on lesson planning and classroom instruction. Teachers could access the project website, which contained sample lesson plans and step-by-step explanations of instructional techniques, and could participate in online chats with the researchers. The project also supported a closed group electronic Listserv to share information, challenges, and successes among participants. Some teachers took more advantage of the coaching and technical support than others.

Teachers in the comparison district did not receive SIOP training. The middle and high school teachers participated in regular district professional development sessions for 2 or 3 days each year of the study. In addition, two ESL teachers led a 1-hour workshop on student diversity and accommodations for ELLs in the classroom at each school. ESL teachers also had training sessions on topics such as creating thematic units; using new content-based ESL textbooks; implementing instructional techniques; and administering and interpreting a new language proficiency test, ACCESS for ELLs (WIDA Consortium, 2006), which measures the academic language of mathematics, science, social studies, and language arts in English.

DATA COLLECTION AND ANALYSIS

As noted earlier, because the professional development program and data collection on student achievement overlapped, we chose to focus analyses on the results of Year 2, by which time most treatment teachers had completed, or had almost completed, their SIOP training.

Classroom Observation Data

Because techniques for sheltered instruction had been implemented, described in the literature, and taught in ESL teacher education programs and staff development workshops for nearly two decades (Crandall, 1993; Stoller, 2004) before we developed the SIOP model, it was reasonable to anticipate that even teachers without SIOP training might incorporate some features of the SIOP model in their lessons. Given our quasi-experimental research design, we wanted to make sure the lesson delivery across the two districts was sufficiently different so that we might ascribe effects on student performance to

the implementation of the SIOP model. Therefore, we measured the extent to which features of the SIOP model were implemented by teachers in both districts. We observed, took field notes, and rated lessons of each teacher twice per year, in the fall and spring. We used the SIOP model's rating scale to score the lessons and assess teachers' level of implementation and fidelity to the intervention. The teachers' scores on individual items and overall percentage scores were recorded in the teacher database.

Based on the observation scores, we determined whether teachers were high, medium, or low implementers and how their level of implementation changed over time. Teachers who scored 50% of the total number of points on the SIOP rating scale or below were considered low implementers; those who scored between 50% and 75% were medium implementers; and those who scored 75% or higher were high implementers.

Student Achievement Data

To examine the impact of the SIOP model on students' English language development, we compiled the students' IPT scores for the Writing, Reading, and Oral Language tests and the Total English proficiency score in a database. The state-mandated IPT tests were administered in March or April each year by district personnel specially trained in the testing procedures.

For both districts, we collected baseline IPT data on all ELLs in the ESL programs from the Spring 2004 administration, before the study began. In subsequent years, in the treatment district we collected IPT scores from those ELLs with at least one SIOP teacher, in Spring 2005 and Spring 2006. Similarly, in the comparison district we collected the IPT scores of ESL students of the participating teachers in Spring 2005 and Spring 2006.

Data Analysis

We examined group means on the IPT for each year, but given the nature of ESL programs, where new students enter and others exit annually, we had a cross-section of students who were not all the same across the years. We report the students' IPT mean proficiency level scores by SIOP treatment and comparison districts in table and graph form to compare the two districts at the outset and to document the changes across the years of the study. Using the Year 2 data, we employed analysis of variance to infer that membership in the group

of students whose teachers received SIOP training can predict some portion of those students' English language achievement as reflected by IPT Writing, Reading, and Oral Language tests. We also calculated the Cohen's *d* effect size. Unfortunately, in these districts only a small number of students participated in the IPT testing for all three administrations, so no longitudinal analyses were undertaken.

To determine whether we should account for school effects in our analyses of the achievement scores, we calculated intraclass correlation (ICC) within and between schools. We wanted to ensure that any increase in scores in treatment schools was due to the treatment and not due to school characteristics. The result of the ICC for all years averaged .03, and for Year 2 particularly it also averaged .03. These results mean that only approximately 3% of the variance in the achievement scores could be attributed to school-level effects. The results also indicate that the schools selected for treatment and comparison were comparable.

RESULTS

Teacher Implementation of the SIOP Model

When we analyzed the data on teacher implementation of SIOP features in the two districts, we found that teachers with SIOP professional development (i.e., treatment teachers) incorporated more features of sheltered instruction than those without (i.e., comparison teachers). At the treatment site, after 1 year of professional development, 56% of Cohort 1 and 74% of Cohort 2 were high implementers of the SIOP model. After 2 years, 71% of Cohort 1 reached a high level. At the comparison site, only 5% of the teachers averaged a high level in the first year; only 17% reached that level by the second year. These results revealed that instruction was different across the two districts and therefore differences in student performance could be related to teacher training and implementation of the SIOP model in the treatment district.

Student Performance on the IPT Tests

Overall, treatment students performed better on the IPT Writing, Reading, and Oral Language tests than comparison students did. Table 4 shows that ELLs in the comparison schools had slightly higher performance on the IPT Writing proficiency level in the baseline year (prior to the SIOP intervention) than the treatment students.

TABLE 4
IPT Writing Mean Proficiency Level Scores for Comparison vs. Treatment Districts

	N	Min	Max	*Mean*	*SD*
Comparison					
Baseline IPT_Writ_PL	176	2.00	5.00	4.16	0.94
Year 1 IPT_Writ_PL	169	1.00	5.00	4.04	1.16
Year 2 IPT_Writ_PL	168	1.00	5.00	4.02*	1.13
Treatment					
Baseline IPT_Writ_PL	386	1.00	5.00	4.06	1.08
Year 1 IPT_Writ_PL	278	1.00	5.00	4.16	0.94
Year 2 IPT_Writ_PL	267	1.00	5.00	4.32*	0.83

Note. *Difference between treatment and comparison groups for that year were significant at $p < .01$.

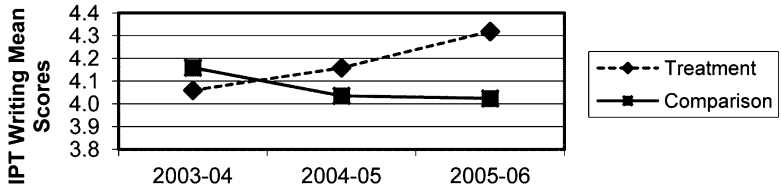

FIGURE 2. IPT writing mean proficiency level scores for ELLs in treatment vs. comparison districts.

However, in Years 1 and 2, treatment students, on average, had higher mean scores than comparison students. By Year 2, this difference was statistically significantly higher, reversing the small achievement gap that had existed at the start.

Figure 2 displays the Writing data graphically and shows the trend that the treatment students' average mean score in writing began below the comparison group in the baseline year, slightly exceeded it in Year 1, and clearly surpassed it by Year 2 of the study, when all but one SIOP training session had been completed in the treatment district.

Table 5 shows a similar trend in Reading, except that treatment students overtook the comparison students only in Year 2. However, the differences in average mean scores were not statistically significant.

Figure 3 illustrates the Reading data graphically and shows the trend that treatment average mean scores in Reading improved each year and ended slightly higher than the comparison district after 2 years, while the comparison district's scores remained flat for the duration of the study.

The two districts were at about the same level of oral language proficiency in the baseline year, but treatment students overtook the comparison students in Year 1 and continued to outperform them in Year 2, as shown in Table 6. In Year 2, the average mean score in the

TABLE 5
IPT Reading Mean Proficiency Level Scores for Comparison vs. Treatment Districts

	N	Min	Max	*Mean*	*SD*
Comparison					
Baseline IPT_Read_PL	188	1.00	5.00	3.95	0.92
Year 1 IPT_Read_PL	169	1.00	5.00	3.98	1.01
Year 2 IPT_Read_PL	168	1.00	5.00	3.97	0.92
Treatment					
Baseline IPT_Read_PL	387	1.00	5.00	3.82	0.92
Year 1 IPT_Read_PL	278	2.00	5.00	3.91	0.88
Year 2 IPT_Read_PL	268	2.00	5.00	4.10	0.83

treatment district was statistically significantly higher than in the comparison district.

The Oral Language data are displayed in Figure 4. The graph shows that although the two districts started out at nearly the same level, the treatment students' average mean scores in oral English improved over the course of 2 years, while the comparison district's remained fairly static.

Total English proficiency level scores show the same trends as the Writing and Oral Language results but have been moderated by the

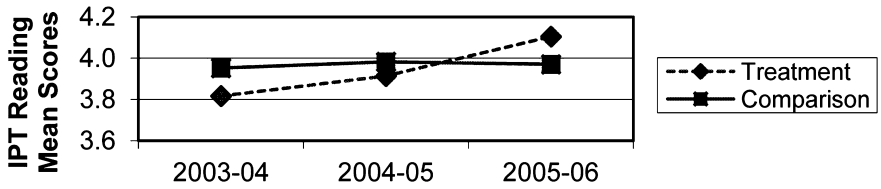

FIGURE 3. IPT reading mean proficiency level scores for ELLs in treatment vs. comparison districts.

TABLE 6
IPT Oral Mean Proficiency Level Scores for Comparison vs. Treatment Districts

	N	Min	Max	*Mean*	*SD*
Comparison					
Baseline IPT_Oral_PL	192	1.00	5.00	3.66	1.28
Year 1 IPT_Oral_PL	169	1.00	5.00	3.65	1.32
Year 2 IPT_Oral_PL	168	1.00	5.00	3.66*	1.33
Treatment					
Baseline IPT_Oral_PL	387	1.00	5.00	3.67	1.37
Year 1 IPT_Oral_PL	278	1.00	5.00	3.76	1.26
Year 2 IPT_Oral_PL	268	1.00	5.00	4.00*	1.06

Note. *Difference between treatment and comparison groups for that year were significant at $p < .01$.

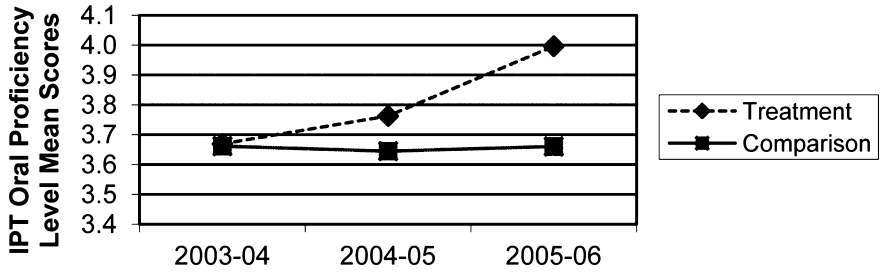

FIGURE 4. IPT oral proficiency level scores for ELLs in treatment vs. comparison districts.

TABLE 7
IPT Total Mean Proficiency Level Scores for Comparison vs. Treatment Districts

	N	Min	Max	*Mean*	*SD*
Comparison					
Baseline IPT_Total_PL	192	1.00	5.00	3.69*	1.04
Year 1 IPT_Total_PL	166	1.00	5.00	3.61	1.23
Year 2 IPT_Total_PL	168	1.00	5.00	3.65*	1.12
Treatment					
Baseline IPT_Total_PL	386	1.00	5.00	3.11*	1.06
Year 1 IPT_Total_PL	278	1.00	5.00	3.70	1.00
Year 2 IPT_Total_PL	267	1.00	5.00	3.88*	0.92

Note. *Difference between treatment and comparison groups for that year were significant at $p < .05$.

lack of strong differences in Reading between the two groups. Nonetheless, Table 7 shows that although comparison district students outperformed treatment students in the baseline year, treatment students surpassed them in Years 1 and 2, showing a statistically significant difference in mean scores in Year 2.

Figure 5 illustrates the Total English proficiency data and shows the comparison district's results overall remained static for the 3 years of the data collection. Despite starting out below the comparison district in the Total English proficiency score in the baseline year, the treatment students' average mean scores slightly exceeded those of the comparison district in Year 1 and then surpassed the comparison district in Year 2.

SIOP as a Predictor of Student Achievement

Analysis of variance of achievement (ANOVA) scores on the SIOP variable provided evidence of SIOP as a predictor of achievement in Writing, Oral Language, and Total English proficiency. Table 8 shows the statistics for each domain test in Year 2. Column 1 is the mean

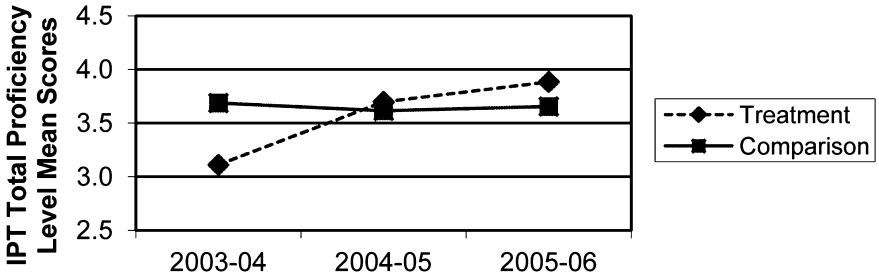

FIGURE 5. IPT total English proficiency level scores for treatment vs. comparison districts.

TABLE 8
IPT Writing, Reading, Oral, and Total Proficiency Level ANOVA and Effect Size Results for Year 2

	Treatment		Comparison					
	Mean	*N*	*Mean*	*N*	β coefficient	*t*	Signifi-cance	Effect size: Cohen's *d*
IPT Writing	4.32*	267	4.02	168	−0.29	−3.12	.00	0.31
IPT Reading	4.10	268	3.97	168	−0.12	−1.58	.12	0.16
IPT Oral	4.00*	268	3.66	168	−0.32	−2.91	.00	0.29
IPT Total	3.88**	267	3.65	168	−0.21	−2.32	.02	0.23

Note. *Difference between treatment and comparison groups for that year were significant at $p < .01$.

**Difference between treatment and comparison groups for that year were significant at $p < .05$.

score for treatment students, and column 3 is the mean score for comparison students. For example, the mean Writing score for SIOP students is 4.32. (Note that this average is also displayed in Table 4, Year 2 treatment district mean.) Comparison students could expect to score lower (−0.29 + 4.32 = 4.02) on the Writing test than treatment students (see Table 4, Year 2 comparison district mean). This increment of −0.29 is about one-quarter proficiency level lower on a scale of 0 to 5 than the expected score for students in treatment classes and is displayed in column 5 of Table 8.

Columns 6 and 7 show the *t*-statistic and significance that were generated in the ANOVA. The Writing and Oral Language differences are significant at the $p < .01$ level, and the Total English difference is significant at the $p < .05$ level in favor of the treatment district. These results of statistical significance indicate that the probability of the differences having been due to chance is less than 5% for the Total English proficiency score and less than 1% for the Writing and Oral Language proficiency scores.

To examine the differences further, we calculated the effect size, a standardized way to report the magnitude of these differences so that

they are comparable regardless of the sample sizes. The final column of Table 8 displays the Cohen's *d* effect size (Cohen, 1988) for the differences in results between the treatment and comparison districts. The result is interpretable as the standard deviation difference between the two groups. In the case of IPT Writing, the treatment group scores were on average almost one third of a standard deviation higher than those of the comparison group (0.31). The treatment scores were better than one fourth of a standard deviation higher for IPT Oral Language (0.29) and close to one fourth for the IPT Total English (0.23).

DISCUSSION

In this study, our main research question was: Will ELLs in one district with teachers who received professional development in the SIOP model show significantly higher achievement in reading, writing, and oral proficiency in English on a standardized measure than ELLs in a comparable district with teachers who had no SIOP model professional development? The significant differences in the average means in favor of the treatment group on Writing, Oral Language, and Total English scores indicate that implementation of the SIOP model had a positive impact on the development of English language proficiency among the ELLs in classes with SIOP-trained teachers. Although the IPT Reading proficiency scores show smaller differences between comparison and treatment groups, the trend favored the treatment schools as well.

It is noteworthy that academic English scores improved significantly, even though most of the ELLs' SIOP instruction took place in mathematics, science, and social studies classes. It strongly supports our hypothesis that both language learning and content learning can take place in subject area classes. The SIOP model, with its attention to language development, had an influence on student English language performance and improved the quality of teaching in the content classes. This is a particularly important finding for contexts where minority language students need to learn the majority language in order to access the core curricular instruction provided through that majority language. Rather than relying on ESL classes as the sole venue for English language development, this result indicates that English language development can take place across the curriculum when teachers use the SIOP model.

There were small to medium effect sizes on IPT Writing, Oral Language, and Total English proficiency for ELLs with SIOP-trained teachers. When we interpret these differences, we must consider that, in general, effect sizes for treatment differences tend to be greatest in the primary grades, with a steady decline as the grades progress (Hill, Bloom, Black, & Lipsey, 2008). Further, as noted in the National

Literacy Panel for Language-Minority Children and Youth report (August & Shanahan, 2006), effect sizes for ELLs are often lower and more variable than those for native-English-speaking students in literacy intervention studies. The ELLs in this study were in Grades 6–12, so an effect size of .31 for the difference in means on the Writing measure and an effect size of .29 for the difference in means on the Oral Language measure show favorable effects of SIOP instruction. The Total English effect (.23) was mitigated by the smaller differences in the results for the Reading test.

While reflecting on possible reasons why the Writing and Oral Language tests yielded better results than Reading, the distinction between receptive and productive language bears consideration. As we have explained, the SIOP model calls on teachers to pay attention to the language in their lessons (how it is used, how it supports learning) and to increase student production of language through oral interaction and writing. ESL teachers already do this to a large extent, but in the typical content classes we have observed in U.S. schools, subject area teachers do not. These teachers talk for more than half of the instructional time, and students listen. If there is a class discussion, it is generally a teacher-led question-and-answer session, during which the teacher asks a question, one or two students provide a brief response, and then the teacher elaborates, corrects, or accepts and moves on to the next question (Cazden, 2001). Some reading, however, does occur through text assignments, literature, and supplemental articles, but there is very little writing in these classes, mostly answering questions in a textbook. Thus, in a typical subject area classroom ELLs primarily use receptive skills.

The SIOP model encourages teachers to change the dynamic and give students more opportunities to talk with other students and to show them how to articulate language functions such as elaboration, justification, and persuasion. We posit that when students learn how to state an opinion or justify a conclusion orally, for example, through the use of sentence starters like "In my opinion …" or "Based on the results, we conclude … ," transfer occurs to their written work.

SIOP teachers build in more writing activities also, ranging from writing sentences using new science vocabulary to filling out an exit slip about something new students learned in class to creating diary entries for a fictional character.

Our initial CREDE study showed significant results for writing, yet it was the only academic skill we assessed. Having a medium positive effect for writing in this subsequent study and a small effect for oral language indicates that implementing the SIOP model can strengthen the productive language skills of ELLs.

The results of the classroom observation data addressed our second question: Do teachers reach high levels of implementation of the SIOP

model during a sustained professional development program after 1 year or after 2 years? We found that 56% of the treatment teachers in Cohort 1 implemented the SIOP model to a high degree after 1 year. By the end of the second year, 71% of these teachers were high implementers. Surprisingly, 74% of the Cohort 2 teachers reached the high implementation level in just 1 year. We suggest that the context of the SIOP model initiative played a role in this variation. The teachers in Cohort 1 participated in a new, district-level initiative in Year 1. The coaching support was limited, and the culture of working in a cross-disciplinary way with a focus on language development was lacking. In contrast, Cohort 2 teachers entered an existing structure and joined teams of teachers who had already experienced success with the initiative. More support was offered from coaches and researchers in the second year, too. Therefore, almost three quarters of the Cohort 2 teachers were able to reach a high level of implementation in 1 year.

Schools, districts, and other institutions that undertake sustained professional development with the SIOP model may expect similar results. Teachers who begin a new initiative may take longer to become high implementers than those who join after the first year. If adding language or content is new to the teachers, time for developing knowledge about that aspect of integrated instruction needs to be planned for, along with learning how to apply new techniques in class. The SIOP model is a complicated framework that calls for significant change in most teachers' lesson planning. It is not one step-by-step technique that can be learned and enacted in a few classroom sessions. Nor is it a set of scripted lesson plans. It is a combination of techniques and requires, on a teacher's part, an awareness of the academic language as used in the subject area along with the knowledge of how to exploit teachable language opportunities when they arise during a class period. Further, teachers need to organize lessons to accommodate more student practice with language—sustained oral interactions and more reading and writing. Given this, having more than half of the Cohort 1 SIOP teachers reach the high level of implementation in 1 year is an affirmative result for this study.

Limitations and Challenges of the Study

Two limitations of this study were the quasi-experimental design and voluntary participation of treatment teachers. The quasi-experimental design was a reasonable next step in the SIOP research program after the CREDE developmental study. This New Jersey study included high school sites, maintained a larger pool of students and teachers, and used state-mandated tests of English language

proficiency as the outcome measures. We recognize that the teachers who agreed to participate in the SIOP training may have been different from the teachers who did not, and those differences may have affected the outcomes, but by Year 2 of the study, when Cohort 2 joined, only 35 students in the ESL programs at the three treatment schools were not included in the student subject pool. In other words, only 35 ELLs did not have at least one SIOP-trained teacher in Year 2. (In Year 1, approximately 70 ELLs were not included.)

A challenge we faced was the selection of the comparison school district. We identified a treatment district easily and received support from the district superintendent for the study. However, given the differences between the middle schools (e.g., one was designated Title I and the other was not) and the intermingled scheduling of teachers and students in the high school, we needed a separate comparison district. In seeking one that matched well on the five factors described earlier, we discovered that persuading a district to act as the comparison site was difficult. We approached several districts before one agreed. Districts declined for several reasons. First, potential districts did not have as much linguistic diversity as the treatment district. They had a preponderance of native Spanish speakers and provided a bilingual program in the middle schools, whereas the treatment district offered an ESL program. Second, we encountered resistance from districts to being labeled a comparison site, despite promised anonymity, support from the New Jersey Department of Education, and an incentive we offered, namely free professional development after the study ended. Third, a few districts were experiencing fatigue after participating in another recent intervention study and were reluctant to take on a new project. As a result, we caution researchers that the challenge of securing a comparison site for quasi-experimental studies, or a control site for randomized experiments, must be factored into future research designs, and new types of incentives may need to be developed.

Directions for Future Research

The design of this study resulted in an overlap of teacher professional development and data collection. Not all teachers had completed their training, nor did they implement the SIOP model to a high degree before student achievement scores were collected. Depending on their class schedule, some students had several SIOP-trained teachers, and one might have been a low implementer while another was a medium or high implementer. Should a similar study be conducted in the future, we propose that more time be given for

teacher implementation before student data collection occurs, and that an analysis of student performance linked to teacher level of implementation be performed. We attempted this type of analysis, but the subsets of students with solely high implementers or solely low implementers were too small for comparison purposes.

We also recommend a future study to collect and analyze student performance data on content exams, with pre- and postmeasures, so that the impact of SIOP instruction on content achievement could be investigated. The New Jersey state tests in mathematics, reading/language arts, and science, unfortunately, did not fit this design. The state gives different tests in these subjects for Grades 6, 7, 8, and 11, so there are no pre- and posttest options for the same group of students. New Jersey also changed the Grades 6 and 7 tests between the first and second year of the intervention, so it was not possible to compare different grade tests that the same students took over the course of 2 years. Moreover, without tests given in Grades 9, 10, or 12, Grade 8 and Grade 11 students do not have 2 successive years of testing. We considered monitoring the performance of all students with additional achievement tests that could be used as pre- and postmeasures, but the district administrators were not in favor of this proposition. There has been considerable testing fatigue in U.S. schools for the past decade due to the increase of federal and state-mandated assessments, so educators did not want to further burden the students.

We also suggest a longitudinal, randomized study of the SIOP model. For example, an experimental study could look at the effects of SIOP instruction on the same cohort of students for 3 to 5 years and compare their achievement to that of a control group of students. This would involve training all the teachers the students might have in advance, ideally 1 year before the cohort is in their classes or grade level so they have the opportunity to be high implementers of the model. However, student mobility would need to be accommodated with a large sample size, because turnover is high in urban-like settings. And a study such as this would need to carefully plan for appropriate pre- and posttest measures to capture growth over time.

Interest in content-based second language teaching is growing worldwide and needs further examination. To date, a significant amount of research on content-based language teaching in international contexts has focused on classroom discourse analysis, student engagement with the content, and teacher perceptions of teaching language or content (Lyster, 2010). We recommend more studies that look at student achievement with the new language or the content material. In places where immersion education is implemented (such as Canada), this research has been conducted, but in other places where content-based language instruction is relatively new or part of

the language majority speakers' schooling, more can be done. One question that might be posed is: What differences, if any, appear in student language achievement when the language integrated with content is the majority language versus the minority language? Another set of related questions might be: How well do students learn the content when it is taught through a second language that they are acquiring? Is there a certain level of proficiency that must be reached in order to learn the content?

CONCLUSION

The results from this study show that the SIOP model offers a promising approach to professional development that can improve the quality of instruction provided to ELLs as well as their academic language achievement in school districts like the ones featured in this study, namely, smaller metropolitan sites of secondary migration and some primary immigration. Although the effect sizes were small to medium, they represent a positive result. Treatment students improved their scores steadily over the 2 years of the SIOP intervention on all four tests, Writing, Reading, Oral Language, and Total English proficiency. The comparison students' scores remained relatively flat during this time. The comparison district results represent what we more customarily find in school districts with ELLs, because each year beginning-level students enter the ESL program, and advanced-level students exit. Therefore, the steady growth of the treatment students is noteworthy.

Although this study took place in the United States, it has value in other contexts as well where second language learners are studying content through the new language. The SIOP model could be applied to English for specific purposes classes, to Dutch as a second language math classes for immigrants in the Netherlands, to French immersion classes in Canada, to CLIL courses, and more. In situations where a systematic model of integrated language and content instruction does not exist, the SIOP model may be used as a framework, or as a starting point for one that would then be modified and tested for these situations. Having a model that teachers use with fidelity facilitates comparative research and evaluation also.

We found that teachers with SIOP model training needed 1 to 2 years of professional development to become high implementers, although almost three quarters of Cohort 2 teachers were high implementers after 1 year. We inferred that this occurred because they joined an existing culture of SIOP implementation and had additional coaching at the high school. Nonetheless, this time frame is important

for administrators and policymakers to note. Changing teacher instructional practice is a long-term endeavor. In this case, it required a series of workshops distributed throughout a school year and additional support through coaching, lesson planning, and technical assistance. Moreover, collaboration among ESL and content teachers is not automatic (Mohan et al., 2001). It requires a vision of cooperative practice, thoughtful coordination, and sustained engagement.

ACKNOWLEDGMENTS

The research study and preliminary analyses were funded by the Carnegie Corporation of New York (Grant #7609) and the Rockefeller Foundation (Grants #2003WC107 and #2005WC144). Additional analyses were funded by the Institute for Education Sciences, U.S. Department of Education, under the auspices of the National Center for Research on the Educational Achievement and Teaching of English Language Learners (Grant #305A050056). The views expressed here are ours and do not necessarily reflect those of the funding institutions. We would like to acknowledge the help of SIOP researchers Dr. Thomas Bauder and Lindsey Hillyard for their work on this study.

THE AUTHORS

Deborah J. Short is a senior research associate with the Center for Applied Linguistics and director of the Academic Language Research & Training consulting firm. Her interests include classroom-based research and professional development in the areas of sheltered instruction and academic literacy.

Carolyn G. Fidelman was a research scientist at the Center for Applied Linguistics but recently took a position at the U.S. Department of Education, National Center for Education Statistics. Her interests include research design and innovative data analysis for educational interventions.

Mohammed Louguit is a research associate in the Language Testing division at the Center for Applied Linguistics. His interests include design and analyses of measures regarding the testing of students learning a new language.

REFERENCES

August, D., & Shanahan, T. (Eds.). (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth.* Mahwah, NJ: Lawrence Erlbaum.

Ballard & Tighe. (2011). *IPT inservice training.* Retrieved from http://www.ballardtighetraining.com/iptinservice/default.html

Cazden, C. (2001). *Classroom discourse: The language of teaching and learning* (2nd ed.). Portsmouth, NH: Heinemann.

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

Crandall, J. A. (1993). Content-centered learning in the United States. *Annual Review of Applied Linguistics, 13*, 111–126.
Echevarria, J., Short, D., & Powers, K. (2006). School reform and standards-based education: An instructional model for English language learners. *Journal of Educational Research, 99*, 195–211. doi:10.3200/JOER.99.4.195-211
Echevarria, J., Vogt, M. E., & Short, D. (2000). *Making content comprehensible for English language learners: The SIOP model*. Needham Heights, MA: Allyn & Bacon.
Echevarria, J., Vogt, M. E., & Short, D. (2004). *Making content comprehensible for English learners: The SIOP model* (2nd ed.). Boston, MA: Pearson/Allyn & Bacon.
Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2006). *Educating English language learners: A synthesis of research evidence*. New York, NY: Cambridge University Press.
Guarino, A. J., Echevarria, J., Short, D., Schick, J., Forbes, S., & Rueda, R. (2001). The Sheltered Instruction Observation Protocol. *Journal of Research in Education, 11*(1), 138–140.
Hill, C. J., Bloom, H. S, Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*, 172–177. doi:10.1111/j.1750-8606.2008.00061.x
Hudec, J., & Short, D. (Producers). (2002a). *Helping English learners succeed: An overview of the SIOP model [Video]*. Washington, DC: Center for Applied Linguistics.
Hudec, J., & Short, D. (Producers). (2002b). *The SIOP model: Sheltered instruction for academic achievement [Video]*. Washington, DC: Center for Applied Linguistics.
Illinois State Board of Education, Assessment Division. (2004). *The Illinois State Assessment: 2004 technical manual*. Retrieved from http://www.isbe.net/assessment/pdfs/isat_tech_2004.pdf
Kindler, A. (2002). *Survey of the states' limited English proficient students and available educational programs and services: 2000–01 summary report*. Washington, DC: National Clearinghouse for English Language Acquisition.
Kober, N., Zabala, D., Chudowsky, N., Chudowsky, V., Gayler, K., & McMurrer, J. (2006). *State high school exit exams: A challenging year*. Washington, DC: Center on Education Policy.
Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam, the Netherlands: John Benjamins.
Lyster, R. (2010, March). *Research perspectives on content-based second language instruction. Colloquium presented at the American Association of Applied Linguistics conference*, Atlanta, GA.
Mohan, B., Leung, C., & Davison, C. (2001). *English as a second language in the mainstream: Teaching learning and identity*. New York, NY: Longman.
Murphy, J., & Stoller, F. L. (2001). Sustained-content language teaching: An emerging definition. *TESOL Journal, 10*(2/3), 3–5.
National Center for Education Statistics. (2009a). *The nation's report card: Mathematics 2009 (NCES 2010-451)*. Washington, DC: Author.
National Center for Education Statistics. (2009b). *The nation's report card: Reading 2009 (NCES 2010-458)*. Washington, DC: Author.
Reardon, S., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal, 46*(3), 853–891. doi:10.3102/0002831209333184
Shavelson, R., & Towne, L. (2001). *Scientific research in education*. Washington, DC: National Academy Press.

Short, D. (2006). Language and content learning and teaching. In B. Spolsky (Ed.), *Encyclopedia of language and linguistics* (2nd ed., Vol. 3, pp. 101–105). Oxford, England: Elsevier.

Short, D., Hudec, J., & Echevarria, J. (2002). *Using the SIOP model: Professional development manual for sheltered instruction.* Washington, DC: Center for Applied Linguistics.

Snow, M., Met, M., & Genesee, F. (1989). A conceptual framework for the integration of language and content in second/foreign language instruction. *TESOL Quarterly, 23*, 201–217. doi:10.2307/3587333

Stoller, F. (2004). Content-based instruction: Perspectives on curriculum planning. *Annual Review of Applied Linguistics, 24*, 261–283. doi:10.1017/S0267190504000108

WIDA Consortium. (2006). *Assessing comprehension and communication in English state-to-state for English language learners (ACCESS for ELLs).* Madison: Board of Regents of the University of Wisconsin System.