Using an Observation Protocol in Bilingual and ESL Classrooms in Institute for Education

Science-Funded Research

Rafael Lara-Alecio, Ph.D.
Professor, Department of Educational Psychology
Director of Bilingual Programs
Texas A&M University
College Station, Texas 77843-4225
a-lara@tamu.edu

Fuhui Tong, Ph.D.
Research Associate, Department of Educational Psychology
Bilingual Programs
Texas A&M University
College Station, Texas 77843-4225
fuhuitong@tamu.edu

Beverly J. Irby, Ed.D.
Professor and Chair, Department of Educational Leadership and Counseling
Sam Houston State University
Huntsville, Texas 77341
irby@shsu.edu

Patricia Mathes, Ph.D.
Texas Instruments Endowed Professor
Institute of Reading Research
School of Education
Southern Methodist University
Dallas, Texas
pmathes@smu.edu

Cindy Guerrero
Ana Quiroz
Kathleen Cox
Doctoral Students
Bilingual Programs
Texas A&M University
College Station, TX

A paper presented at the Annual Meeting of the American Educational Research Association on
Wednesday, April 11, 2007, Chicago, Illinois
Abstract

The purpose of our observational study was twofold. First, by introducing an observation protocol (Transitional Bilingual Observation Protocol [TBOP]) (Lara-Alecio & Parker, 1994), we empirically described teaching behaviors in two types of programs, bilingual and structured English immersion, for kindergarten Spanish-speaking ELLs in a large urban school district. The two program models included an experimental version and a typical practice (control) of each type. Second, we identified variations across the models related to the teachers’ pedagogical approaches. Specifically, two research questions guided our study: (a) What is the time allocation of pedagogical approaches implemented in transitional bilingual education (TBE) and structured English immersion (SEI) language classrooms, as observed by TBOP?, and (b) Do teachers’ pedagogical approaches vary among program models? To describe and compare the characteristics of instruction provided in each condition in our study, teachers were observed providing English language instruction four times across the academic year using the Transitional Bilingual Observation Protocol with 60, 20-second observations using a PDA which increased accuracy in reporting. The data reported in this paper were collected as the final observation of the Kindergarten intervention year in Spring, 2005, totaling 12,898 observations. Our findings indicated that within the same program label of structured English immersion, significant differences were found to be that enhanced classroom teachers were involved in a higher percentage of instruction in (a) intensive English, (b) light and dense cognitive areas, (c) expressive language-related communication, (d) teacher-ask/student-answer type of activity, academic scaffolding and leveled questions, (e) use of English in cognitive area; (f) use of English in expressive language-related communication mode; and (g) academic task rather than social participation task. The same findings also applied to two bilingual programs.
Observation Protocol 3

Using an Observation Protocol in Bilingual and ESL Classrooms in Institute for Education
Science-Funded Research

Introduction

Researchers have affirmed with regard to the variation of classroom characteristics, it is critical to gather observational evidence related to quality instruction that contributes to students’ academic outcomes. (Foorman & Schatschneider, 2003; Foorman et al., 2006; Protheroe, 2002; Waxman & Padrón, 2004). Instructional practices, including language of instruction, are even more critical for those students who are at risk of school failure (Foorman & Schatschneider, 2003; Hill & Flynn, 2006; Uribe, 2004). What is missing from literature, however, is (a) the knowledge base and empirical studies documenting classroom pedagogical occurrences for English language learners (ELLs), (b) the documented quality of instruction by languages of instruction, and (c) the interactions of pedagogy that may produce quality outcomes for such students. Such missing information is supported not only by our own review of literature, but also by recent syntheses from August and Shanahan (2006), Slavin and Cheung (2003), Gersten and Baker (2000), and Thomas and Collier (2003).

Huitt (2003) indicated that complex combinations of various forms of classroom activities takes years for experienced teachers to meet the ideals of instructional practices established by themselves or school districts. This issue becomes even more intricate when the attention is drawn to the pedagogical practices and classroom activities occurring in bilingual and English as a second language (ESL) program models, due to (a) the controversy of in what language to teach and (b) the time-intensive nature and observer-school scheduling alignment issues. Therefore, the purpose of our observational study was twofold. First, by introducing an observation protocol (Transitional Bilingual Observation Protocol [TBOP]) (Lara-Alecio &
Parker, 1994), we empirically have described teaching behaviors in two types of programs, bilingual and structured English immersion, for kindergarten Spanish-speaking ELLs in a large urban school district. The two program models included an experimental version and a typical practice (control) of each type. Second, we identified variations across the models related to the teachers’ pedagogical approaches. Specifically, two research questions guided our study:

1. What is the time allocation of pedagogical approaches implemented in transitional bilingual education (TBE) and structured English immersion (SEI) language classrooms, as observed by TBOP?

2. Do teachers’ pedagogical approaches vary among program models?

Observational Studies

Only a few classroom observation studies have focused on ELLs. Early studies examined instructional events occurring daily in classrooms with ELLs (Breunig, 1998; Brisk, 1991; Escamilla, 1992; Greene, 1997; Heras, 1994; Ramírez, 1992; Strong, 1986). The documentation of classroom instruction initiated as early as in 1980s. Strong (1986) described 20 elementary school teachers’ amount of language instruction (both English [L2] and native language [L1]) and pedagogical activities in TBE and English-only classrooms. The observation was conducted with a coding system to record the proportion of teacher language to silence, as well as the amount of language instruction by bilingual teachers in L1/L2. Regardless of the nature bilingual classroom where more instruction L1 was expected, teachers spent the same amount of time using L2 in both types of classrooms. Similar findings were reported by researches conducted in maintenance two-way Spanish/English bilingual classrooms (Escamilla, 1992) and early-exit TBE (Dolson & Mayer, 1992) that teachers were lacking in their use of L1 (Spanish in both of the studies). Nevertheless Ramírez (1992) concluded that the proportion of English and Spanish
used were consistent with the instructional models defined by the study itself (English-immersion, early-exit, and late-exit TBE).

Recent observational studies largely focus on reading/language arts instruction in elementary classrooms, along with an emphasis on teacher-student interaction and preference to time-sampling approach. Concerned with the paucity of research to involve ELLs in early reading intervention Haager, Gersten, Baker and Graves (2003) designed the English-Language Learner Classroom Observation Instrument for teachers with beginning ELLs. It is a moderate-inference instrument which allows observers to judge the quality of classroom instruction in a pre-determined set of categories including instructional practice, interactive teaching, adaptations for individual differences, English-language development, vocabulary development and phonemic awareness and decoding. Classrooms are observed during the entire reading/language arts period with a minimum of 2.5 hours. Although this instrument is composed of a 4-point Likert rating scale together with a median inter-rater reliability of .74 the authors stated that it will be very useful for research rather than evaluation purpose.

Foorman and Schatschneider (2003) reviewed existing instruments used in observing instructional delivery and proposed a two-dimension measure of time-by-activity to quantify teaching and students engagement, which consists of language codes (the time allotment instructed in either English or Spanish), and 20 content codes (oral language and listening comprehension, vocabulary, phonemic awareness, etc.). One-minute timed observation was conducted throughout the class period and observers rated the overall quality of instruction at the end of each interval. This instrument was adapted by Foorman, Goldenberg, Carlson, Saunders and Pollard-Durodola (2004) in observing three representational models during reading/language arts and/or English language development (ELD) instruction: late-exit TBE; two-way dual
language; and English immersion. This multiyear multisite study totaled 105 classrooms from the Texas and California borders and urban sites with 848 students in kindergarten through second grade. It was observed that teachers in California SEI classrooms instructed exclusively in English, while teachers in Texas SEI classrooms instructed primarily in English with a small portion of Spanish. Irrespective of program model, teachers from California sites allocated more time than their Texas peers in oral language instruction, including oral language/discussion, English language strategies, Spanish language strategies, and vocabulary. With regard to late-exit model, Texas kindergarten teachers were observed to spend 26% of class time in English instruction, which resembled an early-exit rather than late-exit program model. Moreover, at kindergarten level teachers consistently spent a higher percentage of time in word work such as book and print awareness, alphabet letter recognition and reproduction, phonemic awareness, etc. As the grade level progresses an increased proportion of time was devoted to reading comprehension (including discussions of predictable text, previewing to prepare for reading, etc.) in all sites. This may lead to claims that it is problematic if only students’ performances are evaluated without taking into consideration the discrepancy between program labeling and real classroom implementation.

In a later study addressing the issue of whether to implement such a separate instructional block of ELD, the same research team Saunders, Foorman and Carlson (2006) compared observational data of kindergarten classrooms from English immersion, dual-language, transitional bilingual and maintenance bilingual programs. No difference was found in the average percentage of time spent using English during reading and ELD instruction among three types of bilingual program. They concluded that students with separate ELD blocks scored higher than those without an ELD block on English oracy and literacy.
Some observational researchers have targeted classrooms with urban low SES students or African American (Edmonds & Briggs; Foorman & Schatschneider, 2003; Greenwood, Abbott & Tapia, 2003). The Haager et al.’s instrument addressing Spanish speaking ELLs in urban school setting, on the other hand, was completed by rating scale, which may require higher inter-rater reliability for this instrument to be widely used. Furthermore, few of these studies have implemented randomized approach to compare pedagogical differences. In summary, the lack of reliable and valid instruments, and therefore, the empirical evidence masks the actual classroom practice in transitional bilingual /ESL programs with Spanish-speaking ELLs.

Good-quality classroom observations require content validity, interrater reliability, stability over time, and utility. Thirty years ago, Rowley (1978) determined that reliability in observations can “be enhanced by a more representative sampling of occasions, and this is best achieved by using a larger number of shorter observation periods” (p. 172).

Method

This study derived from an on-going five-year federal experimental research project entitled English and Literacy Acquisition (ELLA) (Grant# R305P030032) targeting at approximately 800 Spanish-speaking ELLs receiving services in four program models: (a) typical/control transitional bilingual education (TBE-T), which represents the typical practice in the school district; (b) enhanced/experimental TBE (TBE-E), which represents the intervention of the project; (c) typical/control structure English immersion (SEI-T), and (d) enhanced/experimental SEI (SEI-E) programs.

Sampling and Research Design

The 2x2 factor design (Table 1) depicts the distribution of students at the beginning of 2004 school year (kindergarten). To determine the number of classrooms and students we
conducted a power analysis using Lipsey’s (1990) sample size table (using alpha of .05 and a power criterion of .90) to ensure that our sample size would allow for the detection of educationally relevant, but relatively small effect size differences between our groups. Thus, we began our study with a larger sample that allowed for significant amount of attrition across the four years, and we intended to begin this research with a sample of 288 in each of four conditions (i.e. Enhanced SEI or TBE, Typical SEI or TBE) for a total of 1152 students. Allowing for approximately 35% attrition, we anticipated maintaining approximately 175 students in each condition, for a total sample of 700 students at the end of third grade. In order to accommodate these numbers of students schools were randomly selected with teachers and students nested within each school and program type.

During the kindergarten year of the study, there were 23 total schools randomly assigned to either experimental or control groups. In order to minimize contamination of intervention, a school could only be randomly assigned to receive the intervention or not. Of the 11 schools receiving an enhanced treatment, nine schools received both enhanced SEI and TBE, while the remaining two schools received either enhanced SEI or TBE. Of the 12 schools receiving the typical practice treatment, nine schools received both typical practice SEI and TBE, while the remaining three schools received only SEI. Those three schools were added in November, 2004, due to an unexpectedly low return rate on parental consent forms. In accordance with the state law, randomization was achieved on the basis of schools, instead of individuals. As a result, this project is, by nature, a quasi-experimental study.

**Intervention**

The comprehensive picture of the intervention provided in Project ELLA can be viewed in Figure 1. The interventions provided in TBE-Experimental (TBE-E) are the same ones
Observation Protocol 9

provided in SEI-Experimental (SEI-E). The intervention contains several aspects: (a) extended time for English instruction, (b) structured English intervention, (c) an altered TBE in experimental to be in philosophy and structure a one-way dual language program, (d) on-going professional development, (e) paraprofessionals who receive initial training and ongoing training, (f) District/University leadership and support, (g) two levels/three-tiered approach, and (h) theoretically based upon the Four Dimensional Bilingual Pedagogical Theory (Lara-Alecio & Parker, 1994).

*Altered TBE-E- One-way dual language.* TBE-E is an altered one-way dual language program in that the traditional 50/50 or 90/10 formula for language was altered to be a 70/30 model for Kindergarten. As a one-way dual language program is has the following characteristics: (a) subject matter is taught in the first and/or second language; (b) literacy is developed in the first and second language; (c) classroom consisted of students with same linguistic background; and (d) comprehensible input is provided in English and the second language (Collier, 1992; Kolak Group Inc, 2005).

*Two levels/Three-tiered approach.* Level I intervention was the professional development with the enhanced treatment teachers who received bi-weekly staff development sessions on the following strategies: (a) enhanced instruction via planning, (b) support for student involvement, (c) vocabulary building and fluency, (d) oral language development, (e) literacy development, (f) reading comprehension, (g) parental support and involvement, and (h) reflective practice via portfolio development. Biweekly, the teachers reviewed the STELLA and upcoming lessons of Santillana. They also reviewed the communications games of the paraprofessionals. Paraprofessionals were trained monthly and were provide the program, Communication Games (Lara-Alecio, Irby, & Quiros, 2004), each day for the students. Four on-
site coordinators (full-time employees) monitored and participated in all grant activities. Additionally, a parent involvement program was implemented during the second semester. Teachers were trained in working with parents on literacy development. Two 45-minute training sessions were provided to the parents by the teachers in the enhanced condition. Eighty percent of the parents attended.

Level II of the instructional intervention was provided for the identified ELLA students including the very lowest performing students in experimental classrooms. All interventions were aligned to the Texas standards of English as a second language. The three tiers of instruction were addressed in this level for the students. The first tier was considered the general language arts provided to all students, and in this case, in kindergarten, in TBE the instruction was in Spanish and in the SEI classes, the instruction was provided in English. The second tier was the specific English as a second language instructional time which was increased from 45 minutes in the typical practice classrooms to 75 minutes in the experimental classrooms. The specific instruction for Tier II included oral language development activities structured from Santillana Intensive English- 50 minutes (Ventriglia & Gonzalez, 2000), a research-based curriculum effective in teaching native Spanish speakers English. The program encourages the students to be actively involved in their own learning. Santillana’s curriculum is based on academic content such as math, science and social studies. Question of the Day, Daily Oral Language- 10 minutes (Lakeshore, 1997) in which a chart was used with pre-printed questions that help spark student discussion on a variety of topics. The teacher placed the pre-selected question of the day in a pocket chart along with three to four answer choices. The students addressed the question by answering in complete sentences. The students then placed cards with their names under the selected column that matched their answer choice. This created an instant
graph that the teacher used to make comparisons, generalizations, and ask the students further questions. The pre-selected questions and answer choices were chosen based on the Santillana themes. The final intervention for all students, Story-retelling and higher-order Thinking for English Literacy and Language Acquisition: STELLA (Irby, Lara-Alecio, Mathes, Rodriguez, & Quiroz, 2004), was designed to be delivered in a 15-minute lesson during the ESL block. Teachers received scripts prior to the week of story introduction allowing them time to practice. During classroom instruction, they introduced one book a week accompanied by a script which included three vocabulary words per book, a pre-selected ESL strategy aligned to the story, and a set of different leveled questions identified as easy, moderate, and difficult. The activities included a mixture of dramatization and music to allow students to use their motor skills. Finally, in Tier III, instruction was for the lowest-functioning students as identified by teachers via students’ classroom functionality and was composed of communication games (20 additional minutes) delivered by highly trained paraprofessionals. This student intervention included small group oracy and literacy instruction with all students in the classroom. Moreover, students received intensive English tutorials by trained paraprofessionals delivered in small groups to the lowest achieving students.

**Ongoing professional development.** Teachers received bi-weekly staff development sessions on the following strategies: (a) enhanced instruction via planning, (b) support for student involvement, (c) vocabulary building and fluency, (d) oral language development, (e) literacy development, (f) reading comprehension, and (g) parental support and involvement. The ESL strategies were selected for kindergarten teacher training from Herrell and Jordan’s (2004) 50 most effective strategies including: Academic Language Scaffolding-Visual and Modeled Talk, Bridging, Communication Games, Dramatization and Scripting, Interactive Read Aloud,
Leveled Questioning, Manipulative and Realia Strategy, Preview/Review, Partner Work and Tutoring, Sorting Activity, Think Aloud, and Total Physical Response with Music and Movement. Biweekly, the teachers reviewed the STELLA and upcoming lessons of Santillana Intensive English Program. They also reviewed the communications games of the paraprofessionals. Paraprofessionals were trained monthly and were provide the program, Communication Games, each day for the students. Teachers were trained in working with parents on literacy development. Four on-site coordinators (full-time employees) monitored and participated in all grant activities.

*Theoretical Framework, Transitional Bilingual Pedagogical Theory and Instrument,*

*Transitional Bilingual Observation Protocol – TBOP*

To describe and compare the characteristics of instruction provided in each condition in this study, teachers were observed providing English language instruction four times across the academic year using the Transitional Bilingual Observation Protocol with 60, 20-second observations using a PDA which increased accuracy in reporting (Figure 2). Rowley (1978) indicated that reliability could be improved by “a more representative sampling of occasions, and this is best achieved by using a larger number of shorter observation periods” (p. 172). All observers were trained and inter-rater reliability was initially taken at .89 with a final reliability established at .98. Two other reliability checks were established during the year.

As reported in more detail by Lara-Alecio and Parker (1994), the Four Dimensional Transitional Bilingual Pedagogical (TBP) Theory originally was developed to identify the interactions of four major instructional dimensions within bilingual classrooms; however, since that time, the Bilingual Observation Protocol that was developed and validated from the Theory (Bruce et al, 1997; Bruenig, 1998), has been applied successfully to evaluation research in, of
course, transitional classrooms, but also, dual language and SEI classrooms with Kappa values ranging from .65 to .98. This four-dimensional Theory, in Figure 3, allowed us in this study to assess the occurrences of language of instruction, language of response in relation to communication mode, cognitive response levels, and instructional activity structures within the classroom within subject matter.

*Language Content.* This domain derives from Cummin’s (1986) influential language acquisition theory distinguishing Basic Interpersonal Communications Skills (BICS) and Cognitive-Academic Language Proficiency (CALP) language competencies. While the BICS and CALP distinction was initially useful, the main limitations (Wiley, 1996) of this simple dichotomy are that it has obscured all classroom communication on a continuum between BICS and CALP, and has discouraged examination of student progress in this vast “middle area.” The Lara-Alecio and Parker Theory reformulates BICS and CALP as malleable levels of discourse, rather than as fixed or long-term abilities. The Theory includes four levels of language content: (1) Social Routines (e.g., social exchanges and conversation), (2) Classroom Routines (e.g., repetitive school-related tasks), (3) Light Cognitive Content (e.g., discussing community news), and (4) Dense Cognitive Content (e.g., entailing conceptually demanding, specialized vocabulary; critical thinking).

*Language of Instruction.* The Model’s second domain, the “Language of Instruction,” presents four progressive uses of native [(L1) (Spanish)] and second [(L2) (English)] language in the classroom: (a) content presented in L1 (Spanish), (b) L1 (Spanish) introduces L2 (English), (c) L2 (English) supported and clarified by L1 (Spanish), and (d) content presented in L2 (English). This dimension acknowledges the concept of “transition” (as in “transitional bilingual”), and affirms the importance of the content areas as rich sources of language input for
ELL students (Cummins, 1986) and as vehicles for language learning (Krashen, 1985). Language of instruction usually refers to the teacher’s use of language. However, it also may refer to the reading text used, or the language used by students in cooperative learning groups. Although the model depicts transition of language, the model can also be used, singling out either L1 or L2, as would be in an all-Spanish or all-English classroom (or as applies to other languages); thus, there is applicability to other types of programs other than transitional bilingual.

Communication Mode. This domain distinguishes two receptive models (Aural, Reading) and two expressive language modes (Verbal, Writing). Cummins’ (1986) “reciprocal interaction model” and the “context-specific” model support the practice of multiple modalities for second language acquisition. These modalities (especially Reading, Writing, and Verbal Expression) also are important curriculum skill areas. Their differentiation within the TBP Theory indicates that English facility may not be unitary, but may vary by communication mode.

Activity Structures. Activity structures are teacher-structured, stable, recurring learning situations, each with its own expectations for teacher and student communication (Brophy & Evertson, 1978; Doyle, 1981). Communication that is expected and fostered in one activity structure may be inappropriate and discouraged in a second. Our traditional pedagogical emphasis on “the lesson” with objectives, curriculum content, and assignments, unfortunately ignores “activity structures.” Influenced by Vygotsky’s notion of Zone of Proximal Development (Cole & Griffin, 1983), classroom ethnographers similarly describe the “structure of events,” each type of structure with its own opportunities, implied values and expectations for student participation (Erickson, 1982). Activity structures are operationally defined in the Lara-Alecio and Parker Theory as combinations of (a) type of teacher behavior (e.g. directing, leading, evaluating, observing), and (b) the expectation for student responding (e.g. listening, performing,
discussing, asking questions, answering questions, cooperative learning). A few classroom activity structures (e.g. time spent disciplining, transitions between classes) are considered non-academic. Most classroom activity structures are defined by combinations of two activities, signifying the main teacher behavior plus the primary student expected behavior. Thus when a teacher mainly lectures or presents information, and students are mainly expected to listen, the activity structure is identified as lecture/listen (Lec/Lis).

To further understand instructional pattern, observations of ESL strategies and curriculum areas were also included in the PDA observation tool.

Data Collection and Analysis

Sixty, twenty-second timed observations were conducted throughout the class period in each classroom and coded into a PDA system, which was then uploaded into the computer. The data reported in this paper were collected as the final observation of the Kindergarten intervention year in Spring, 2005, totaling 12,898 observations. Chi-square test of homogeneity of proportion was employed to determine the differences of cross-classification among the four instructional deliveries on each domain proposed in TBO theory by using SPSS for Windows, version 14.0. In the case when the null hypothesis of homogeneity or equal proportion \( H_0: \pi_1 = \pi_2 = \pi_3 = \pi_4 \) was rejected, a post hoc pair-wise 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 \( \alpha \) level (Type I error), chi-square test of homogeneity maintains \( \alpha \) at a constant level throughout the significant tests (Cox & Key, 1993). Cramer’s \( V \) was also reported as type of effect size in our study (Rea & Parker, 1992).
Results

Results were presented based on the four domains proposed by the Lara-Alecio and Parker Pedagogical Theory together with ESL strategies, as well as the interaction among these domains. Seventy-four percent of the observations were conducted during ESL block, with another 16% conducted during reading/language arts. Table 2 lists overall chi-square of homogeneity results for the main and interaction effect of domains tested.

Language of Instruction

Chi-square was significant at $\alpha = .05$, with a Cramer’s V of .2, indicating that the association between the variables is at the edge of moderate in strength and worth noting. Particularly to the interest of this study, post hoc pair wise comparison was performed in the domain of teachers’ L1 and L2 (Table 3) for chi-square difference test ($\chi^2 (0.05, 1) = 3.84$). The experimental teachers were observed less frequently speaking in Spanish during the ESL teaching time (SEI-E: .26%; TBE-E: .14%) than the control classrooms teachers (SEI-T: 8.5%; TBE-T: 14.40%). To the contrary, the SEI-E (97.3%) and the TBE-E (98.3%) teachers were observed speaking in English at a higher rate during their ESL instructional time than the SEI-T (86.1%) and the TBE-T (75.4%) teachers. All differences are statistically significant at $p = .05$ ($\Delta \chi^2 (1) > 3.84$) except for the use of Spanish between teachers in SEI-E and TBE-E classrooms. Interesting findings resulted from the examination of the language used by students. Figure 4 illustrates the pattern of students’ language use corresponding to their teachers’ instruction.

Language of Content

Chi-square test was significant at $\alpha = .05$, with a Cramer’s V of .2, indicating a moderate association. Figure 5 demonstrates a higher percentage of social language and academic routines
observed in typical practice classrooms than in enhanced classrooms while higher percentage of light cognitive and dense cognitive content observed in enhanced classrooms than in typical practice classrooms. All post hoc pair wise comparisons within each sub-domain yielded statistically significant differences except for the light cognitive content area where the percentage of time allocation was equivalent between teachers in two typical practice classrooms.

Communication Mode

Chi-square test was significant at $\alpha = .05$, with a Cramer’s V of .14, indicating that the association between the variables is at the edge of weak to moderate in strength. Eighteen different language communication modes were observed as they were elicited from students by their teachers (Figure 6). The most frequent single mode was verbal with a higher percentage observed in enhanced classrooms than typical classrooms. The most frequent combination of modes observed was aural-verbal (au-ver) with it more frequently observed in the enhanced classrooms (SEI-E: 43.4%; TBE: 41.8%) as opposed to the typical practice classrooms (SEI-T: 28.3%; TBE-T: 40%). Listening (aural) was observed more frequently in typical practice classrooms (SEI-T: 19.9; TBE-T: 20.4%) than in enhanced classrooms (SEI-E: 12.4%; TBE-E: 18.7%). Although with low frequencies, writing and reading were observed more often in typical practice classrooms (SEI-T: 5.12%; TBE-T: 7.73%) than in enhanced classrooms (SEI-E: .82%; TBE-E: .49%). Post hoc pair wise comparisons were presented in Table 3 on selected sub-domains with frequent occurrences.

Activity Structures

Chi-square test was significant at $\alpha = .05$, with a Cramer’s V of .2, indicating a moderate strong association. The activity structure most frequently observed was “ask/answer.” This was
with greater frequency in the enhanced classrooms (SEI-E: 54.4%; TBE-E: 43.1%) as opposed to the typical practice classrooms (SEI-T: 26.6%; TBE-T: 37.3%). The next most frequently observed activity structure was “lead/perform” with greater frequency in the enhanced classrooms than typical classrooms. Nonacademic activities transition was occurred more frequently in typical classrooms as opposed to enhanced classrooms. Student interaction, although with low occurrence, was observed more frequently in enhanced classrooms (SEI-E: 4.4%; TBE-E: 3.9%) than typical classrooms (SEI-T: 2.2%, TBE-E: 3.2%), although the difference was not statistically significant. This is illustrated in Figure 7.

**ESL Strategies**

Overall chi-square test was significant at $\alpha = .05$, with a moderate strong association. Figure 8 depicts five most frequently implemented ESL strategies among four program models. Academic scaffolding and leveled questions were observed with a higher percentage of instruction in enhanced classrooms than typical classrooms. The post hoc test difference chi-square values calculated from contingency table was statistically significant. Non-academic strategies were used more often in typical than enhanced classrooms (Figure 8). Other strategies used more frequently in enhanced classrooms over the typical practice classrooms were: manipulatives and realia, partner work, preview/review, think aloud, total physical response, and dramatization.

**Interaction: Language of Instruction by Content — L2**

In light of the fact that most of the observations (90%) were conducted during ESL instruction and reading/language arts, the use of English was of particular concern. The chi-square test of the interaction effect between language of instruction and language content was significant at $\alpha = .05$, with a weak association. It can be implied from Figure 9 that when
instructing in L2, typical classroom teachers spent more time teaching social and academic routine while enhanced classroom teachers, following the intervention objectives, spent more time in cognitive areas. All post hoc chi-square difference were larger than 3.84 (Table 3). During English instruction, all teachers were observed spending approximately half of the instructional time on light cognitive area, with more time in enhanced classrooms (SEI-E: 55.4%; TBE-E: 58.5%) than typical classrooms (SEI-T: 47.3%; TBE-T: 48.8%).

Interaction: Language of Instruction by Mode — L2

Likewise, the use of English was of primary interest when examining the interaction between language of instruction and communication mode. Chi-square test was significant at \( \alpha = .05 \), with a Cramer’s V of .25, indicating that the association between the variables is at the edge of moderate in strength and worth noting. Consistent with patterns observed in the domain of communication mode, the two most frequently observed modes during L2 instruction were aural-verbal and verbal, with a higher percentage in enhanced classrooms as opposed to typical classrooms. Non-academic mode was observed occurring more frequently in typical classrooms than enhanced classrooms. Additionally, more frequency in reading-related modes was observed in typical practice classrooms (Figure 10).

Discussion and Conclusion

Researcher have devoted to the relative effects of different English learning programs to identify best practices that promote ELLs’ language and academic achievement. Nevertheless, it is also worth looking at the potential value of within-program comparisons (Saunders, Foorman, & Carlson, 2006). Our findings indicated that within the same program label of structured English immersion, significant differences were found that enhanced classroom teachers were involved in a higher percentage of instruction in (a) intensive English, (b) light and dense
cognitive areas, (c) expressive language-related communication, (d) teacher-ask/student-answer type of activity, academic scaffolding and leveled questions, (e) use of English in cognitive area; (f) use of English in expressive language-related communication mode; and (g) academic task rather than social participation task. The same findings also apply to two bilingual programs. It was observed that control teachers spent a significant proportion of time in teaching academic routine and light cognitive in L1, while no L1 instruction occurred in cognitive area in experimental classrooms, instead, experimental teachers spent more time using L1 in social and academic routine. To further examine this seemingly implausible case, we found that L1 instruction only occurred .1% of the total observation period (8 out of 2940 and 3 out of 2819 observations, respectively), indicating that during ESL intervention, teachers were following the outlined objectives in teaching English, with an emphasis on teaching cognitively demanding content and vocabulary development. This is even more critical because students’ use of language mirrored their teachers’ instruction. Therefore, if the teachers purposefully choose academic-related language in L2, students will respond in a similar manner, which in turn, provides opportunities for students to engage in the target language learning. The CALP in L2, which, according to Cummins’ theory, takes as long as seven years to acquire, determines subsequent L2 academic performance for ELLs, as Collier (1987) suggested, “Language proficiency required for school tasks can incorporate the whole range of skill… but it is especially in school that students need to develop context-reduced and cognitively demanding aspects of language in order to function successfully in the classroom”(pp. 618-619). An early intervention with intensive English instruction in cognitively demanding areas can accelerate such acquisition.
In addition, when L2 was the language of instruction by teachers, listening-speaking and speaking occurred in a significant portion of communication modes in both enhanced classrooms. However, in all receptive language-related mode (reading and listening), typical classroom teachers spent more time than the enhanced classroom teachers did. There appeared to be more ‘teacher talk’ and ‘student talk’ in typical SEI and TBE classrooms. Carrasquillo and Rodriguez (1995) indicated that most talk comes from the teacher and not from the students: they stated that students should have time to talk. Although the teacher-ask/student-answer type of activity was more frequently observed in enhanced classrooms, which may lead one to speculate that students were passive listeners or respondents to teachers’ questions, it was in fact in alignment with the academic scaffolding and leveled questions as most frequently implemented ESL strategies which were more effective during asking/answering activity. Moreover, in elementary classrooms as early as in kindergarten, this type of activity is reasonably predominant (Hill & Flynn, 2006). Our findings indicated that students in enhanced classroom were more frequently encouraged to engage in interaction with peers than were those in typical practice classrooms. This was also suggested by Short (1993) and Carrasquillo and Rodriguez (1996) to employ collaborative or cooperative learning activities in classroom with ELLs.

To summarize, the data reflected the expected pedagogy as established by the project and the effectiveness of the project’s Level I intervention – professional development. Our study reflected reliable judgments of how teachers allocate instructional time, language, content and modes, as well as the interaction among these modes in the enhanced and typical structured English immersion and transitional bilingual classrooms. Finally, the TBOP is a flexible and comprehensive classroom observational instrument that can be used in different educational
settings. It is also a reliable research and evaluation tool that provides a good picture of teachers’ pedagogical patterns and their interaction with students (ELLs).
Observation Protocol 23

References


Author Note

This article is part of an extensive on-going research grant funded by the U.S. Department of Education, Institution of Education Sciences (IES), Project English Language and Literacy Acquisition (ELLA), R305P030032.
Table 1

_Distribution of Students At the Beginning of 2004 School Year - Kindergarten_

<table>
<thead>
<tr>
<th></th>
<th>SEI</th>
<th>TBE</th>
<th>Total n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enhanced</strong></td>
<td>Schools: 10</td>
<td>Schools: 10</td>
<td>Schools: 10</td>
</tr>
<tr>
<td></td>
<td>Classrooms: 13</td>
<td>Classrooms: 17</td>
<td>Classrooms: 30</td>
</tr>
<tr>
<td></td>
<td>Students: 1174</td>
<td>Students: 290</td>
<td>Students: 464</td>
</tr>
<tr>
<td><strong>Typical Practice</strong></td>
<td>Schools: 12</td>
<td>Schools: 9</td>
<td>Schools: 19</td>
</tr>
<tr>
<td></td>
<td>Classrooms: 19</td>
<td>Classrooms: 11</td>
<td>Classrooms: 30</td>
</tr>
<tr>
<td></td>
<td>Students: 176</td>
<td>Students: 182</td>
<td>Students: 358</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>Schools: 22</td>
<td>Schools: 19</td>
<td>Schools: 60</td>
</tr>
<tr>
<td></td>
<td>Classrooms: 32</td>
<td>Classroom: 28</td>
<td>Classrooms: 60</td>
</tr>
<tr>
<td></td>
<td>Students: 350</td>
<td>Students: 472</td>
<td>Students: 822</td>
</tr>
</tbody>
</table>
### Chi-square Test Statistics for All Domains by TBOP

<table>
<thead>
<tr>
<th>Domain</th>
<th>Chi-square(^a)</th>
<th>Degree of freedom</th>
<th>Cramer's V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language of Instruction</td>
<td>1507.396</td>
<td>12</td>
<td>0.2</td>
</tr>
<tr>
<td>Language Content</td>
<td>735.393</td>
<td>9</td>
<td>0.14</td>
</tr>
<tr>
<td>Communication Mode</td>
<td>2203.34</td>
<td>51</td>
<td>0.24</td>
</tr>
<tr>
<td>Activity Structures</td>
<td>1426.235</td>
<td>60</td>
<td>0.2</td>
</tr>
<tr>
<td>ESL Strategies</td>
<td>2254.569</td>
<td>39</td>
<td>0.24</td>
</tr>
<tr>
<td>Language of Instruction * Language Content – L2</td>
<td>535.716</td>
<td>9</td>
<td>0.13</td>
</tr>
<tr>
<td>Language of Instruction * Mode – L2</td>
<td>2097.577</td>
<td>51</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Note. \(^a\)p < .001 for all test statistics.
Table 3

Chi-square Statistics Calculated from Contingency Table in all Domains among Four Program Models

<table>
<thead>
<tr>
<th>Domain</th>
<th>Sub-domain</th>
<th>SEI-E n=2940</th>
<th>SEI-T n=4319</th>
<th>TBE-E n=2819</th>
<th>TBE-T n=2820</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language of Instruction</td>
<td>L1</td>
<td>162.6</td>
<td>40.6</td>
<td>165.0</td>
<td>323.1</td>
</tr>
<tr>
<td></td>
<td>L2</td>
<td>23.0</td>
<td>4.1</td>
<td>27.5</td>
<td>58.2</td>
</tr>
<tr>
<td>Language Content</td>
<td>light cognitive</td>
<td>10.01</td>
<td>14.90</td>
<td>30.97</td>
<td>16.14</td>
</tr>
<tr>
<td></td>
<td>dense cognitive</td>
<td>12.90</td>
<td>113.02</td>
<td>91.67</td>
<td>0.01</td>
</tr>
<tr>
<td>Communication Mode</td>
<td>au-ver</td>
<td>28.68</td>
<td>91.84</td>
<td>15.91</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td>verbal</td>
<td>63.59</td>
<td>226.58</td>
<td>216.34</td>
<td>17.80</td>
</tr>
<tr>
<td></td>
<td>aural</td>
<td>33.56</td>
<td>23.61</td>
<td>22.57</td>
<td>21.63</td>
</tr>
<tr>
<td></td>
<td>re-ver</td>
<td>14.16</td>
<td>15.99</td>
<td>6.06</td>
<td>40.06</td>
</tr>
<tr>
<td></td>
<td>au-re</td>
<td>12.61</td>
<td>22.31</td>
<td>64.21</td>
<td>2.13</td>
</tr>
<tr>
<td>Language content in L2</td>
<td>social routine</td>
<td>1.90</td>
<td>7.08</td>
<td>51.34</td>
<td>39.19</td>
</tr>
<tr>
<td></td>
<td>academic routine</td>
<td>20.38</td>
<td>110.75</td>
<td>87.67</td>
<td>4.05</td>
</tr>
<tr>
<td></td>
<td>light cognitive</td>
<td>5.29</td>
<td>17.93</td>
<td>20.18</td>
<td>4.82</td>
</tr>
<tr>
<td></td>
<td>dense cognitive</td>
<td>8.84</td>
<td>78.97</td>
<td>74.85</td>
<td>2.48</td>
</tr>
<tr>
<td>Communication mode in L2</td>
<td>au-ver</td>
<td>27.10</td>
<td>94.88</td>
<td>11.77</td>
<td>8.54</td>
</tr>
<tr>
<td></td>
<td>verbal</td>
<td>37.45</td>
<td>207.00</td>
<td>162.48</td>
<td>6.90</td>
</tr>
<tr>
<td></td>
<td>aural</td>
<td>24.85</td>
<td>17.39</td>
<td>15.62</td>
<td>22.85</td>
</tr>
<tr>
<td></td>
<td>au-re</td>
<td>12.85</td>
<td>39.63</td>
<td>64.82</td>
<td>10.82</td>
</tr>
<tr>
<td></td>
<td>reading</td>
<td>66.61</td>
<td>230.45</td>
<td>35.62</td>
<td>14.37</td>
</tr>
</tbody>
</table>

Note. $\chi^2 (.05, 1) = 3.84.$
Figure 1. Project ELLA model.
Figure 2. TBOP on PDA
Figure 3. Four dimensional transitional bilingual pedagogical theory model (Parker & Lara-Alecio, 1994)
Figure 4. Language of instruction and language of students among four program models
Figure 5. Language of content among four program models.
Figure 6. Five most frequently observed communication modes among four program models.
Figure 7. Six most frequently observed activity structures among four program models.
Figure 8. Selected ESL strategies implemented among four program models.
<table>
<thead>
<tr>
<th></th>
<th>social routine</th>
<th>academic routine</th>
<th>light cognitive</th>
<th>dense cognitive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SEI-E</strong></td>
<td>2.3%</td>
<td>26.6%</td>
<td>55.4%</td>
<td>15.6%</td>
</tr>
<tr>
<td><strong>SEI-T</strong></td>
<td>3.5%</td>
<td>41.0%</td>
<td>47.3%</td>
<td>8.2%</td>
</tr>
<tr>
<td><strong>TBE-E</strong></td>
<td>0.5%</td>
<td>21.4%</td>
<td>58.5%</td>
<td>19.6%</td>
</tr>
<tr>
<td><strong>TBE-T</strong></td>
<td>5.0%</td>
<td>33.8%</td>
<td>48.8%</td>
<td>12.3%</td>
</tr>
</tbody>
</table>

*Figure 9. Cross domain interaction: Language of instruction by content – L2*
Figure 10. Cross domain interaction: Language of instruction by mode – L2