Student Performance in a Content-Based Second Language Program

Cynthia McPhail
Rochester Public Schools

The format of a bilingual education program at an urban school in Western New York changed from language arts classes, to enrichment classes in science and social studies which were conducted in the children's second languages. Goals of the change were to improve the children's performance in social studies, science, and language (English and Spanish). This ex post facto study utilized ANOVA and ANCOVA analyses to determine if differences in English, Spanish, social studies, and science performance were statistically significant. For both Spanish and English dominant students it was found that English scores remained the same. Mean Spanish scores were not significantly different in all analyses, but Spanish growth rates for the students who participated in the program either increased or stayed the same, while growth rates of those who did not participate decreased; there was an overall improvement in both social studies and science. Findings must be interpreted keeping in mind various confounding variables. Discussion includes recommendations for addressing these variables in future studies.

Introduction

The format of a bilingual program at an urban school in Western New York changed from a focus on language arts to one in which the curriculum emphasized enrichment classes for science and social studies which were conducted in the children's second language. The goals of the change were to improve the children's performance in social studies, science and language (English and Spanish). The purpose of this study was to determine if these goals had been met.

Review of the Literature

The idea that children can acquire a second language through carefully planned input and interaction in their second language when incorporated into content area lessons such as mathematics, social studies and science has received a great deal of support (Brinton, Snow & Wesche, 1989; Chamot & O'Malley, 1987; Christian, Spanos, Crandall, Simich-Dudgeon, & Willets, 1990; Garcia 1990; Mohan 1986; Slavin, 1986; Snow, 1991). Language development approaches in which teachers integrate English language development with regular basic skills instruction are effective (Milk, 1985). Language ceases to be taught in isolation when learning occurs in the communicative environment of the content classroom (Mohan, 1986). Indeed, for most learners, second language acquisition "will take place only to the extent that those learners are exposed to and engaged in contextually-rich, genuine, meaningful communication in that language" (Taylor, 1983, p. 70).

Content area instruction provides the basis for language learning in that it contains an inherent feature of naturalistic learning, real meaning (Snow, Met and Genesee, 1989). It "can provide both a motivational and a cognitive basis for language learning," providing "a primary motivational incentive for language learning insofar as it is interesting and of some value to the learner and therefore worth learning" (Snow, Met and Genesee, 1989 p. 202).

Cummins (1981) points out that language demands can be described as being *context reduced* or *context embedded*. *Context embedded* language is rich in contextual cues such as concrete objects, gestures, visual aids and facial expressions. *Context reduced* language reduces the amount of these cues so that the ability to understand is dependent entirely on listening comprehension. Context reduced language is generally more difficult than context embedded language for the second language learner. Thus, favorable subjects for second language instruction are determined by the amount of practical activity, or the amount that language is used in connection with visible situations which illustrate meaning (Stoddart and Stoddart, 1968). The criterion for a productive natural communication situation for

beginning students is that the material be visually demonstrable (Dulay and Burt, 1975), or context embedded. Chamot and O'Malley (1987) recommend utilizing experiences in a second language related to science, mathematics, and social studies, beginning with science instruction since it can focus on a discovery approach "that will capitalize on experiential learning opportunities that provide both contextual support and language development" (p. 231). According to Mohan (1986), both students and teachers find it easier to communicate through "experiential approaches, such as demonstrations and practical tasks. Moreover, less advanced students will find it easier to communicate with each other in practical tasks; group work like building a model often produces a great deal of communicative interaction" (p. 117-8).

Since an important factor for ease of acquisition is the amount to which the student is familiar with the content of a class (Stoddart and Stoddart, 1968), Chamot and O'Malley (1987) suggest that teachers build on students' previous knowledge by providing then with experiences to develop new concepts and expand upon previous ones.

While amount of input can play a role in second language acquisition, utility of input is not solely a function of quantity (Harklau, 1994). Effective input also has a communicative purpose that emphasizes the content of the message rather that its grammatical form (Krashen, 1981a, 1981b; Tikunoff, 1985; Wong-Fillmore, 1985). Second language classes that focus exclusively on linguistic features are missing a potentially powerful element of instruction.

Student attitudes also affect second language acquisition. Harklau (1994) found that students tended to devalue ESL instruction compared to mainstream classes. This may be because:

> Students' efforts to achieve will be lowest when they think they can't master the material, when they think that mastering the material is not very important, very fun or very useful; or when they do not know what they need to do in order to move onto the next step of task mastery (Eccles and Wigfield, 1985, p. 187).

Also, "By providing students opportunities to use language in meaningful contexts-studying the academic subject matter while they develop language proficiency- teachers create an ideal learning environment" (Short, 1993, p. 629).

Given the support for content area second language classes, this researcher was intrigued by a principal's interest in examining his students' performance in such a program. He had already piloted the program with a set of third grade classes, and had implemented the program in both sets of bilingual fourth and fifth grade classes for a total of two years at the time this article was written. Since the desire to measure achievement was expressed after the program had been in place, readily available data were used to conduct an ex post facto, empirical examination of the students' achievement.

Sample

This study examined the fourth and fifth grade performance of 112 students attending a school in a predominately Hispanic urban area of a mid-sized Western New York city. Fifty-five percent of the school population of over 500 students from grades 3-8 is considered limited English proficient (LEP) as determined by Language Assessment Scale measures and standardized test scores such as the Comprehensive Test of Basic Skills (CTBS). The racial/ethnic composition of the student body is 0.6% Native American or Native Alaskan, 0.4% Asian or Pacific Islander, 15% African American of not Hispanic origin, 75% Hispanic, and 9% White of not Hispanic origin. Eighty-eight percent of the students qualify for free/reduced price lunches. Seven percent of the students receive special education services, but neither these students nor those who were retained in the fourth or fifth grades were included in this study. Average classroom size is approximately 25 students with each grade level consisting of four classes. Three classes are bilingual and labeled English dominant, Spanish dominant, and Spanish/English transitional. The fourth class is for English dominant children from the area who do not wish to participate in a bilingual program. Students are placed in classes by English language proficiency. Only

students in the bilingual classes were included in this study.

Each bilingual class was instructed by a full time teacher who was accompanied by a paraprofessional 50% of the time (on alternating days). All teachers had Masters degrees and NYS certification (probationary or permanent) or the equivalent. Years of experience ranged from three to fifteen with a balance of veteran teachers in both grades.

Treatment

Language dominance was determined through reports of the home language survey and through Language Assessment Scale (LAS) testing. Both English and Spanish dominant children who attend the school receive instruction in their second languages. Before the second language class program change to content based instruction, Spanish Language Arts instruction took a variety of formats such as literacy development through basal readers, oral skills and vocabulary development through a picture card program, and whole language classes. Generally, these classes were held in a large group/entire class format. Spanish dominant children received English reading instruction with basal readers, and English as a Second Language (ESL) instruction in a small group, pull-out setting. While following a language skills-focused, city-wide ESL curriculum, instruction would also often be coordinated with the students' English basal readers, focusing on skills that were weak, as well as occasionally incorporating topics of interest to the students such as holidays and some content area lessons. All regular content area classes were conducted in the children's dominant language.

In 1992, the two-way coordinator met with teachers to discuss possibilities for changing the second language class format. According to a copy of the 1992 proposal for funding, it was determined that:

> Second language acquisition will be enhanced through instructional activities in the content areas. The program will implement an integrated curriculum based on thematic units.

> These units will integrate the program's language and content subject objectives. Teachers continue the gradual increase of the second language during routine activities and use of vocabulary related to content areas such as math, social studies and science.

As described in the program proposal, students were to receive native language instruction in social studies and science for 35 minutes daily, math for 50 minutes daily, native language reading and language arts for a total of 2 hours daily, and second language class for 30 minutes daily, in addition to classes such as art, music and physical education on a rotating basis.

A pilot study of this program was implemented for the 1992-1993 school year in the third grade. Grades 4-6 followed suit in the 1993-1994 school year.

Method

Variables

This study measured program effectiveness, the independent variable, through examining a series of dependent variables. Because this study was conducted ex post facto, analyses were limited to previously existing measures of performance that had been archived at the school. Available measures for all of the years included in this study were the following: English Language Assessment Scale (LAS) measures for the limited English Proficient (LEP) students, CTBS Reading Total Scores for all students, Spanish Assessment of Basic Education (SABE) Reading Total Scores for all students, and raw scores on the city-wide final exams in social studies and science for all students. A description of each of these instruments follows.

Language Assessment Scale (LAS). This study utilized the LAS as a covariate only. The LAS is an individually administered test designed to measure English oral language proficiency. Level I of the test is designed for grades K - 5, and is the level that was

administered to the sample. The LAS is composed of seven subtests which assess phonology, auditory discrimination, the lexicon, syntax, sentence comprehension, writing (optional) and pragmatics (optional). Neither of the optional portions were administered.

The test is normed so that monolingual first graders pass virtually all of the items in Level I. According to a review of the instrument, the manual includes "an impressive amount of evidence on the reliabilities and validities of the LAS" (Haber, 1985).

Scores on each of the subtests are tallied, totaled and transformed into an ordinal scale ranging from 1-5 with levels 1 and 2 representing non-speakers, 3 representing a limited speaker, and 4 and 5 representing a proficient speaker. In order to differentiate between fluent non-native speakers and native speakers of English in the statistical analyses in this study, native speakers were coded as 6. This test was administered every year to non-native speakers of English in the sample. Spanish LAS scores were not available.

Comprehensive Test of Basic Skills (CTBS). All students in the school must take the CTBS regardless of their native language. The CTBS is a nationally normed, standardized, multiple choice test widely used for testing reading and math skills in English. Students in the sample completed both the English Reading Comprehension and Vocabulary subsections, however, only the combined score on these subsections (the Total Reading Score) was reported in the student files. Thus, analyses included only the Total Reading Scores which were reported in percentiles. This test is administered to the children in the spring of every year.

Spanish Assessment of Basic Education (SABE). All students in the program must take the SABE regardless of their dominant language. The SABE is a series of norm-referenced tests which have been designed to allow comparison to the CTBS test, although they are conducted in Spanish. Sample students completed the Reading Comprehension and Vocabulary subsections of the tests. These scores, like the CTBS scores, were reported in the student files only as a Total Reading Score which was reported in percentiles. This test is administered to the children in the spring of every year.

City-Wide Social Studies and Science Exams. City-wide final exams are dependent on the year's curriculum and are redesigned every year. The exams are developed by the City-wide head of each department in conjunction with a staff of teachers and test developers. The tests are designed to measure elements of the New York State Board of Regents' curricula for social studies and science which are provided to the teachers, thus providing for a level of standardization across the city. The exams are administered in June each school year. Students are allowed to take the exams in the language of their choice because the exams are translated by a team of bilinguals from the Language Assessment Center in the city.

Procedure

Fourth and fifth grade scores on the dependent variables (CTBS, SABE, and city-wide exams) dating from 1991-1995 were available for the three sets of students in this study.

In order to verify if there was a difference between no-program vs. program performance, in addition to comparing the averages of each single year's results, student growth was compared. Growth from fourth grade to fifth grade in each of the dependent variables was calculated for the (A) 1991-1992/1992-1993, (B) 1992-1993/ 1993-1994, and (C) 1993-1994/1994-1995 school years (here-on-in, years will be identified by the calendar year of the spring semester and groups will be identified by the aforementioned letters: A , B, or C). Group A represents the no-program comparison group since these students did not participate in the new program in either fourth or fifth grade. It would have been valuable to obtain growth rates for other no-program years, however, much of these data were unavailable. Group B represents a group that received no-program instruction in fourth grade, and program instruction in fifth. Group C represents a group that received program instruction in third, fourth and fifth grade since they had participated in the third grade pilot of the program. A visual display of the groups is in Figure 1.

Placement in the three groups was not random. Therefore, there is a possibility of confounding variables such as non-equivalent

groups, and the danger of a teacher effect due to the relatively small sample. Before comparing the groups' growth rates then, it was necessary to determine if there was an overall equivalence. Distribution of each group's English language ability in fourth grade, as determined by the LAS, was compared. It was found that there was an unequal distribution, with group A having the most English native speakers, B having the most low level English speakers as shown in Table 1 (LAS scores of 1-2 being considered non-speakers and 3 being a limited speaker), and C having the most students scoring a 5. Spanish proficiency data were not available.

Group	1991/92	1992/93	1993/94	1994/95
A n=37	4th grade *No program*	5th grade *No program*		
B n=40		4th grade *No program*	5th grade **Program**	
C n=35		3rd grade **Pilot**	4th grade **Program**	5th grade **Program**

Figure 1. Display of years and grades of participation

LAS	1992 (A)			1993 (B)			1994 (C)		
	Freq	Tot. %	val. %	Freq	Tot. %	val. %	Freq	Tot. %	val.%
1	0	0.0	0.0	4	10.0	11.1	1	2.9	3.0
2	1	2.8	3.0	2	5.0	5.6	0	0.0	0.0
3	2	5.4	6.1	2	5.0	5.6	3	8.6	9.1
4	2	5.4	6.1	5	12.5	13.9	2	5.7	6.1
5	12	32.4	36.4	14	35.0	38.9	20	57.1	60.6
6	16	43.2	48.5	9	22.5	25.0	7	20.0	21.2
0	4	10.8	-	4	10.0	-	2	5.7	-

Table 1. LAS Level Distributions Within Each Fourth Grade

Results

CTBS

Because one would expect some degree of correlation between a measure of language proficiency and achievement when achievement is tested through a written test, it was not surprising to find that the correlation between Fourth grade English LAS level and Fourth grade CTBS performance was 0.553, significant at $p < .01$ in a two-tailed test. This indicates that the unequal distribution of highly correlated English LAS scores and CTBS scores must be taken into account before comparing the groups' growth. An ANCOVA using LAS level as a covariate was calculated. Because of the covariate, any differences in fourth grade CTBS scores between the three groups could be deemed significant over and above any differences caused by the LAS level distributions. After English LAS level was entered as a covariate, the ANCOVA revealed no significant differences between the fourth grade performance of the three groups on the CTBS, $F(2,87) = 1.229$, $p = .298$. This allows us to compare the three groups' growth in CTBS scores since the groups' fourth grade performances, or starting points, can be considered equivalent once accounting for unequal English language level distribution.

An ANCOVA for amount of growth was calculated using LAS level as a covariate. It was found that there were no significant differences in the amount of growth in each of the LAS levels, $F(1,81) = .001$, $p = .987$. Likewise, there was no significant difference in the amount of growth in each of the three groups, $F(2,81) = 1.908$, $p = .155$. Thus there is no significant difference in growth in CTBS performance.

Non-native English speaking children's scores were examined. It was logical to also examine these scores separately since one of the goals of the content based second language program was to improve non-native speakers' English proficiency. Another ANCOVA using only LAS levels 1-5 as a covariate was conducted. This, as well, showed no significant differences in amount of CTBS growth between the three groups, $F(2,58) = 1.633$, $p = .204$.

SABE

Analysis of the SABE scores proved to be more complicated. Fourth grade English LAS scores were correlated with Fourth grade SABE performance at a -0.3199 level ($p < .01$). Thus, when comparing the groups to determine overall equivalency in fourth grade, we can expect the unequal language level distributions to have an effect. An ANCOVA of fourth grade SABE performance in each of the groups revealed a significant difference, $F(2,87) = 5.210$, $p < .007$. This tells us that the covariate English LAS score did not account for all of the differences in performance between the three groups. This, most likely, is because English LAS score is not the best variable to use as a covariate. A level of Spanish proficiency would be much more appropriate. Unfortunately, Spanish LAS data were not available for all of the students, especially those who were English dominant and acquiring Spanish.

Consequently, it is useful here to look at descriptive data. Table 2 illustrates mean scores on the SABE in fourth and fifth grade, as well as mean amount of growth.

Group A had the highest mean in fourth grade, and group B had the lowest. Group A's scores decreased the most from 4th to 5th grade, while group B's increased the most, and group C remained essentially the same. An ANCOVA of the differences in SABE growth using LAS scores as a covariate indicates a significant difference in growth between groups, $F(2,84) = 5.142$, $p < .008$.

Group	4th Grade	5th Grade	Growth[1] (5th-4th)
A	42.71	33.16	-9.55
B	25.81	42.9	17.09
C	35.83	34.81	-1.02

[1]Raw 5th grade means minus raw 4th grade means will not equal the raw growth means because of attrition. Not all students had scores for both 4th and 5th grade.

Table 2. SABE Means

It is interesting to note that an ANCOVA of fifth grade scores using LAS level as a covariate showed no significant differences between the groups' SABE scores in fifth grade, F (2,89) = .112, p = .894. Thus, the groups' status changed from having unequal performance on the SABE in fourth grade, to having essentially equal performance in fifth.

Although the sample population is small, it is nevertheless worthwhile to examine the English native speakers' Spanish performance. An ANOVA of SABE growth scores (i.e., fifth grade score minus the fourth grade) was significant at $p < .001$, $F(2,19) = 20.339$. The means were: A) -.42 (n=12), B) 28.0 (n=5), and C) 3.00 (n=5), which approximate the trends in rate of growth demonstrated within the group at large.

City-Wide Final Exams

Final exams are not the same from year to year. It is inappropriate to compute the difference between raw scores as a measure of growth since the difficulty of each test may not be the same. In order to measure growth, it was necessary to compute the sample's Z-scores using the means and standard deviations from the entire population of students taking the tests in the city each year, convert the Z-scores to percentiles, and then conduct comparisons.

The city does not archive raw scores, only frequencies of scores occurring in five point intervals ranging from 70 to 100. Scores below 70 are only recorded as failing. In order to calculate the means and standard deviations for the population's performance on each year's test, data bases of the entire city's performance on each of the 12 tests were created. Since the raw scores were unavailable, midpoints of each interval were used. Means and standard deviations of the city's performance on each of the 12 exams were used to calculate Z scores for the sample's performance. Since the number of students in the city taking the final exams was quite large, (3,114 to 3,371), it is appropriate to assume a standard distribution of scores. Consequently, the sample's Z-scores were converted to percentiles using a standard normal distribution chart.

Since scores below 70 were not included in the city's archives, they also were not included in calculating the sample's z-scores. However, failure rates for the city and the school were calculated. These manipulations provided two indicators of test performance: percentile means of passing scores, and percent of failure, or failure rates for the school and the city.

Social Studies

A one-way ANOVA of the percentile means (Table 3) for the three groups' passing scores indicated that there was a significant difference for fourth grade $F(2,84) = 4.365$, $p < .016$. An examination of the means shows group C outscoring the other two. This could, perhaps, be attributed to group C's having participated in the pilot program in the third grade.

An ANOVA comparison of the three groups' percentile means from passing scores in Social Studies in fifth grade was not significant, $F\ (2,92) = 1.766$, $p = .177$. The failure rates for each of the three groups in 4th grade as compared to the city are presented in Table 4.

Group	4th Grade	5th Grade
A	29.21 (n=31)	45.51 (n=34)
B	30.82 (n=30)	33.59 (n=30)
C	48.67 (n=26)	34.99 (n=31)

Table 3. Social Studies Means

	Fourth Grade		Fifth Grade	
	Group	City	Group	City
A	35.1	14.7	16.2	16.8
B	29.7	17.8	27.5	25.6
C	21.9	15.7	8.8	21.1

Table 4. Social Studies Failure Rate

Each year the school has had a fourth grade failure rate which was higher than the city's. However, the trend over the three years was for the school to improve, when the city did not. Group A and B's fifth grade failure rates were more or less equivalent to that of the city but group C's performance is notably better.

Science

Fourth grade percentile means (Table 5) for passing scores were compared with an ANOVA which was significant at $p < .001$, F (2,80) = 12.670. From the means and an examination of the groups' failure rates (Table 6), we can see that group B performed more poorly than the other two groups in fourth grade. An ANOVA of fifth grade percentile means of passing scores was not significant, F (2,85) = 3.000, $p = .055$).

An examination of the data shows that group A's percentile means of passing scores decreased, and failure rate increased from fourth to fifth grade. Percentile means of passing scores for B and C either

Group	4th Grade	5th Grade
A	48.92 (n=32)	33.44 (n=28)
B	18.34 (n=24)	32.39 (n=31)
C	50.19 (n=27)	48.56 (n=29)

Table 5. Science Means

	Fourth Grade		Fifth Grade	
	Group	City	Group	City
A	13.5	18.3	30.6	20.1
B	59.5	29.1	30.8	24.7
C	21.9	20.1	20.6	33.7

Table 6. Science Failure Rates

increased or stayed essentially the same. Failure rate for both groups B and C improved, with group C achieving the only occurrence of a fifth grade science exam failure rate lower than that of the city.

Teacher Interviews

After concluding the empirical analyses of the data, teacher interviews were conducted in order to gain insight into the implementation of the program. Comments will be included in the cautions and discussion sections.

Findings

Discussion

CTBS. The results indicate that the program has had no effect on students' growth as demonstrated by CTBS scores, regardless of their LAS level.

SABE. Although fourth grade means were different, there was no significant difference in the three groups' fifth grade means. This would appear to signify no major positive impact on growth that can be attributed to the program. However, an examination of the groups' growth rates indicates that group A, the group who never participated in the program, decreased from 4th to 5th grade. The other two groups either improved or maintained their Spanish levels. This is certainly preferable to the loss demonstrated by group A.

Social Studies. The failure rate improved from fourth to fifth grade for all three groups, and fifth grade passing score means were not significantly different from each other. However, group C's rapid improvement in failure rate with no significant decrease in mean of passing scores as compared to the other groups, indicates an apparent improvement in social studies.

Science. Group B changed from being significantly lower in fourth grade, to not being significantly different in fifth grade, as

well as having an improved failure rate; and group C essentially maintained the same percentile mean of passing scores while achieving the only occurrence of a failure rate lower than the city. This indicates overall improvement in science.

Cautions

A study with no confounding variables is very rare. This study is no exception. Conclusions must be interpreted with an understanding of these weaknesses.

The first weakness is that the differences could be due to a teacher effect. There was a range of teaching experience from three to fifteen years with some changes in faculty. While there was a balance of veteran teachers in both grades, and the fourth grade teachers remained relatively constant with only one teacher change during the four years, the fifth grade teachers changed considerably from year to year, with only one of the teachers participating in all years of the study.

There may be problems with the measures of the dependent variables. The LAS provides a measure of oral proficiency, but Haber (1985) cautions that if the test is to be used for a more fine-grained analysis of language proficiency, it is inadequate due to serious practical difficulties of administration and scoring as well as serious theoretical and methodological difficulties in test construction. Haber does, however, complement the test on its wealth of widely and carefully normed data and states that it is one of the finest measures of oral proficiency known to her.

While the CTBS and SABE are recognized for their validity and reliability in terms of assessing reading ability, in this study they have been used as measures of language proficiency. According to Bachman (1991) language proficiency is multicomponential.

Therefore, a Reading Total Score is not an accurate picture of language proficiency. A further caveat is that since students reach the 50th NCE on standardized language arts tests two to four years earlier than they reach the 50th NCE on reading tests, reading tests are better predictors of student academic performance in the second language at the secondary and postsecondary levels rather than elementary as they were used in this study (Collier, 1989).

According to Saville-Troike (1984) "vocabulary knowledge is the single most important area of second language competence when learning content through that language is the dependent variable" (p. 199). Nevertheless, science and social studies vocabulary may be different than general reading. The vocabulary scores from the CTBS and SABE tests may misrepresent students' vocabulary knowledge. "A test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned" (Hughes, 1989, p. 22). Teachers who use content-based language instruction "recognize the need for additional assessment instruments more commensurate with the academic demands of the mainstream curricula" (Short, 1993, p. 632). "Assessment is the weak link in the integrated language and content approach" (Short, 1993, p. 653).

Because the City archives its final exam data in intervals, rather than raw data, calculations of the means and standard deviations of each science and social studies test are only rough estimates.

By using the midpoint of each interval, the variance within each interval is lost. Thus, the Z-scores are also only rough estimates. When making any conclusions from the percentiles, we must keep in mind that roughly estimated Z-scores were used in the conversion. There was, however, no other empirical manner in which to conduct analyses to answer the question of whether social studies and science performance had improved.

An additional confounding variable is that the program's implementation was not the same from year to year. According to a teacher participant, the first year was more experimental. Only in the second year did the teachers feel more able to collaborate and share ideas. The program, thus, evolved from one year to the next. This change could affect the data. We should also note that the program had been implemented for a relatively short time. It is possible that significant changes would not arise until after a greater period. In addition, the amount of second language instruction time was quite small.

The final, and perhaps the most damaging confounding variable is that a second program change occurred concurrently with the second

language/content class program change. A teacher participant explained that before the 1993/1994 school year, science and social studies were not taught daily. Instead, either science or social studies would be taught daily for approximately an hour until a unit was complete. In this way, the students received a full hour of instruction in social studies or science, but the two classes were never taught together in the same day. This schedule change has profound effects on this study. According to the teacher:

> Both social studies and science (scores have improved), but I don't think that it's from our program. I think that social studies and science scores have gone up because now we teach it every day. Where it wasn't really done every day the whole time before, now we give 45 minutes of social studies every day, 45 minutes of science every day. Where before we were flip flopping science and social studies, and I think that is what was hurting us a lot...but now that we teach it every day, I think that's the key. I don't think it's the content in Spanish.

There is no way to separate the effects of these simultaneous program changes. We have no way of knowing if the improved content area performance was a result of one change or another. This last finding illustrates the importance of conducting studies of both process *and* product in second language pedagogy as recommended by Lambert (1991), Long (1991), and Spada (1987). Too often, researchers fail to examine what is actually going on in the classrooms concerned in a study, and after the conclusion, there is no way of knowing whether teachers adhered to a method or approach.

Conclusions

Given the numerous confounding variables potentially affecting the results of this study, all conclusions must be viewed with healthy skepticism.

The program had no effect on English proficiency as measured by the CTBS. It lead to maintenance of, or improvement in Spanish proficiency as measured by the SABE. Examinations of failure rates and sample percentile means indicate that performance in both science and social studies has improved, but there is no way to tell if this outcome was the result of the program change, or the change in daily schedules. The program does not appear to be detrimental to student performance in any of the four dependent variables. However, it has not achieved the goal of improved performance in all of them.

Recommendations for other studies are: 1) Analyze the process as well as the product of a program in order to examine the extent of its implementation; 2) Establish potential measures of dependent variables at the onset of a program, not post hoc; 3) Use the most appropriate instruments to measure the variables, and insure that they are administered correctly; 4) Collect data on variables that are known to correlate with the dependent variables to use as covariates; 5) Archive data in the most accurate fashion possible; 6) Document the implementation of any program changes. These six recommendations would have decreased the impact of a majority of the confounding variables in this study. Keeping these suggestions in mind will help to improve future research projects.

References

Bachman, L. F. (1991, Winter). What does language testing have to offer? *TESOL Quarterly, 25*, 671-705.

Brinton, D. S., Snow, M. A., & Wesche, M. B. (1989). *Content-based second language instruction*. New York: Newbury House Publishers.

Chamot, A. U., & O'Malley, J. M. (1987). The cognitive academic language learning approach: A bridge to the mainstream. *TESOL Quarterly, 21*, 227-249.

Christian, D., Spanos, G., Crandall, J., Simich-Dudgeon, C. & Willetts, K. (1990). Combining language and content for second language students. In Padilla, A., Fairchild, H., & Valadez, C. (Eds.) *Bilingual education: Issues and strategies* (pp. 141-156). Newbury Park, CA: Sage Publications.

Collier, V. P. (1989). How Long? A synthesis of research on academic achievement in a second language. *TESOL Quarterly, 23, (3),* 509-531.

Cummins, J. (1981). The Role of primary language development in promoting educational success for language minority students. In California State Department of Education, Office of Bilingual Bicultural Education, *Schooling and language minority students: A theoretical framework* (pp. 3-49). Los Angeles, CA: Evaluation, Dissemination and Assessment Center.

Dulay, H. C. and Burt, M. K. (1975). Creative construction in second language learning and teaching. In M. Burt, & H. C. Dulay (Eds.), *On TESOL '75: New directions in second language learning and bilingual education* (pp. 21-32). Washington, DC: Teachers of English to Speakers of Other Languages.

Eccles, J., and Wigfield, A. (1985). Teacher expectations and student motivation. In Duseck, J. (Ed.) *Teacher expectancies* (pp. 185-217). Hillsdale, NJ: Lawrence Erlbaum Associate, Publishers.

Garcia, E. (1990). Instructional discourse in "effective" Hispanic classrooms. In Jacobson, R. & Faltis, D. (Eds.), *Language distribution issues in bilingual schooling* (pp. 104-132). Philadelphia, PA: Multilingual Matters Ltd.

Haber, L. (1985). Review of language assessment scales. In James V. Mitchell, (Ed.) *The ninth mental measurement yearbook* (pp. 808-811). Lincoln, Nebraska: Buros Institute of Mental Measurements of the University of Nebraska.

Harklau, L. (1994). ESL versus mainstream classes: contrasting L2 learning environments. *TESOL Quarterly, 28, (2),* 241-271.

Hughes, A. (1989). *Testing for language teachers.* New York, NY: Cambridge University Press.

Krashen, S. (1981a). Bilingual education and second language acquisition theory. In California State Department of Education, Office of Bilingual Bicultural Education, *Schooling and language minority students: a theoretical framework* (pp. 51-79). Los Angeles, CA: Evaluation, Dissemination and Assessment Center.

Krashen, S. (1981b). *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Lambert, W. E. (1991) Pros, cons, and limits to quantitative approaches in foreign language acquisition and research. In Freed, B. (Ed.) *Foreign language acquisition research and the classroom* (pp. 321-337). Lexington, MA: Heath.

Long, M. (1991) The design and psycholinguistic motivation of research on foreign language learning. In Freed, B. (Ed.) *Foreign language acquisition research and the classroom* (pp. 309-320). Lexington, MA: Heath.

Milk, R. D. (1985). The changing role of ESL in bilingual education. *TESOL Quarterly, 19, (4)*, 657-672.

Mohan, B. A. (1986). *Language and content*. Reading, Massachusetts: Addison-Wesley Publishing Company.

Saville-Troike, M. (1984, Winter). What really matters in second language learning for academic achievement? *TESOL Quarterly, 18 (2)*,199-220.

Short, D. J.(1993, Winter). Assessing integrated language and content instruction. *TESOL Quarterly, 27, 4*, 627-656.

Slavin, R. E. (1986, Summer). Learning together. *American Educator*, 6-13.

Snow, M. A. (1991). Teaching language through content. In Celce-Murcia, M. (Ed.) *Teaching English as a second or foreign language* (pp. 315-327). New York, NY: Newbury House.

Snow, M. A., Met, M., and Genesee, F. (1989). A conceptual framework for the integration of language and content in second/ foreign language instruction. *TESOL Quarterly, 23, (2)*, 201-217.

Spada, N. (1987). The relationship between instructional differences and learning outcomes: A process-product study of communicative language teaching. *Applied Linguistics, 8*, 137-161.

Stoddart J., and Stoddart, F., (1968). *The teaching of English to immigrant children*. London: University of London Press.

Taylor, B. (1983). Teaching ESL: Incorporating a communicative, student-centered component. *TESOL Quarterly 17, 1*, 69-88.

Tikunoff, W. J. (1985). Applying significant bilingual instructional features in the classroom. *Part C Bilingual Education Research Series*. National Clearinghouse for Bilingual Education. (ERIC Document Reproduction Service No. ED 338 108).

Wong-Fillmore, L. (1985). When does teacher talk work as input? In S. M. Gass and C. G. Madden (Eds.), *Input in second language acquisition*, (pp. 17-50). Rowley, MA: Newbury House.