A Synthesis of Studies Examining Long-Term Language Minority Student Data on Academic Achievement

Virginia P. Collier
Graduate School of Education
George Mason University

Abstract

This article provides a research synthesis of studies that have examined language-minority students' academic achievement over a period of four or more years, for a comparison with the longitudinal findings on student academic achievement reported in the Ramírez study. One program variable is the focus of this synthesis -- the use of a minority language for instructional purposes. Some of the limitations to long-term research are discussed, followed by a summary of results from a variety of language-minority studies conducted in the United States on two-way bilingual education, late-exit bilingual education, early-exit bilingual education, and programs with no first language support. Implications for program effectiveness decisions for language-minority students are provided.

Introduction

Program effectiveness is currently a hot topic in education. Evaluations of effectiveness of school programs are typically conducted when federal or state funding is involved, or a school board wants to examine a pilot program, or a school system is under pressure to improve student achievement. Almost all stakeholders in these evaluations want instant or short-term answers. They want to know, in one or two years, what the results are.

However, educators rarely get a true picture of student progress with one or two years of student performance data. If a program is new, and staff and students are excited about the innovation, the Hawthorne effect (Borg & Gall 1983) can influence early student gains, but may wear off in succeeding years. A quick look at student data for any program in the early grades (K-2) almost always appears to show reasonable student gains, but for the few programs that allow a continuing look at succeeding years, upper elementary and secondary achievement generally is less impressive. Normal growth curves on standardized tests typically flatten out as

students move up in age and grade level, as the curriculum becomes academically and cognitively more complex.

We language-minority educators must face this reality. Quick fixes do not work in education, neither in programs nor in program evaluation. When we examine short-term results, we are getting an inaccurate picture of true student achievement.

This article provides a research synthesis of studies that have examined language-minority students' academic achievement over a period of four or more years, for a comparison with the longitudinal findings on student academic achievement reported in the Ramírez study (Ramírez, Pasta, Yuen, Ramey & Billings 1991). While there are many potential program, instructional, and student background variables that may influence language-minority student performance, this review must, for limitations of space, focus on examination of only one of these variables. The chosen program variable for this synthesis is the main focus of the research questions of the Ramírez study--the use of a minority language for instructional purposes and its influence on language-minority student achievement. Some of the limitations to long-term research are discussed, followed by a summary of some of the results from a variety of language-minority studies conducted in the United States. Implications for program effectiveness decisions for language-minority students are provided.

Some limitations to long-term studies

A wide range of studies in the United States and in other countries have examined the academic achievement of language-minority students across several years of participation in a given program. Each study that has been published has some serious limitations in the study design, but increasingly useful studies are being conducted and are worthy of examination. Among the most common problems are limits that have to do with the difficulty of controlling for all variables that affect analysis of the results. For example, student background variables are very likely to have an influence on student achievement; yet many studies do not control for crucial variables because the student population is so varied and permission is not given to collect this data for reasons of privacy. As an example of the difficulty of controlling the program treatment variables, the term "bilingual program" is used to refer to hundreds of variations in actual implementation. Rarely do studies include measures of process variables that examine details of the program treatment from classroom to classroom, as did the Ramírez study.

This is because very little research funding is available to conduct studies of the magnitude of the data collection effort exerted in the Ramírez study.

Another limitation to most studies is that they present cross-sectional rather than longitudinal student data. Instead of comparing the same students across time as would be done in a longitudinal analysis such as the Ramírez study, most studies must report cross-sectional analyses of all language-minority students participating at each grade level in order to have a sufficient number of cases to present generalizable results. Even in the Ramírez study, the decision to present only longitudinal analyses limited the results to four years of data, which is a minimal length of time for a long-term examination of program effects. Although the designers of this study had considered collecting student data over a period of more than four years, by the fifth year of implementation, the number of students remaining was too few for program results to be generalizable. Perhaps in the long run, cross-sectional analyses are a necessary compromise. Following the same students across time in a longitudinal analysis generally yields higher student results than those of a cross-sectional analysis. Thus a cross-sectional picture probably provides a more conservative estimate of how a group of students may fare in a given program.

A third limitation of long-term studies in general are the limits of the measures used to evaluate students' academic achievement. In order to compare the performance of one group of students with that of another group, standardized tests are the only measure providing a normative standard. Yet a standardized test in English typically underestimates limited-English-proficient students' actual academic achievement, and it may take several years of exposure to English before the test becomes a more valid measure for these students. For students studying academically in their second language, current standardized measures for content assessment still generally rely on multiple-choice items that measure only one language skill, reading, and do not measure listening, speaking, or writing skills. Standardized performance assessment being developed at the present time may address this concern, but establishing meaningful normative standards will be difficult.

Given these limitations, it is still very worthwhile to examine data reported in long-term studies to see what insights they can provide for patterns across time in language-minority student academic achievement. The studies reviewed here have been chosen

because they are available to the public in some published form, and they have reported findings in sufficient detail to allow for comparisons of student achievement across a period of at least four years.

Findings of the Ramírez study

The program variable to be examined in this review is the one that has caused the greatest controversy in the field of language-minority education in the United States: the use of a minority language for instructional purposes. This was the bottom-line purpose of the Ramírez study--to compare language-minority students' performance in a variety of program treatments, in which the major difference between programs was the amount of instruction conducted in students' first language (L1). The research questions of the Ramírez study were designed to answer policy questions concerning U.S. federal funding for language-minority programs. What level of use of students' L1 for instructional purposes works best?

The three program types examined in the Ramírez study were very carefully matched, so that the chief distinguishing characteristic between programs was the amount of L1 instruction provided to students, with other program and instructional variables controlled. This study went far beyond any previous study in analysis of program treatment variables, through observation of each classroom and extensive data coding focused on teaching style, levels of classroom interaction, levels of L1 and L2 use, and student academic engagement. The three programs were found to be very similar in general classroom instructional patterns, student-teacher ratio, and teacher experience and preparation. All three programs required teachers to be bilingual and taught language through academic content. The main program difference was the amount of time that the instruction was conducted in Spanish, the primary language of all language-minority students in this study.

In all three programs observed in the Ramírez study, in the initial stages, students were allowed to speak in Spanish; however, teachers' use of Spanish varied significantly between the three models. In the structured immersion strategy classrooms, all instruction was in English; while in early-exit bilingual classes, Spanish was used for a short period of time (part of the day in Grades K-2) for limited L1 literacy and cognitive development, concurrently developed with L2 literacy; and in late-exit bilingual

classes, Spanish literacy was developed before introducing English literacy, and once English literacy was initiated, approximately 50 percent of the instructional day was devoted to cognitive/academic development in each language through Grade 6. More simply stated, structured immersion strategy provided no Spanish instructional support; early-exit provided very limited Spanish support for only the early grades; and late-exit provided substantial, balanced Spanish and English support through Grade 6. What were the results in students' long-term academic achievement?

In the Ramírez study, longitudinal data on students' academic achievement were collected for Grades K-3 for all three programs. Data was also collected from an additional student cohort in late-exit programs for Grades 3-6, to provide a measure of 'their performance in L2 in the upper elementary grades. Since this program model introduces L2 literacy instruction after a strong base in L1 literacy is established, students in a late-exit program in third grade would not have had as much exposure to L2. Therefore, a standardized test in L2 given in third grade would underestimate their actual cognitive/academic development. Additional data was thus needed from Grades 4-6 in late-exit programs to measure a number of years of exposure to L2 comparable to the immersion strategy and early-exit students, who received more L2 instruction in the lower elementary grades.

The analyses of the student achievement data in the Ramírez study are admirably done, using very complex and current statistical procedures that are daunting to interpret. The explanations and interpretations of the data in the complete final report are thorough, and therefore lengthy, but very useful. However, the summaries presented to the media and the public by the U.S. Department of Education have not accurately presented the actual findings of the study. For the policy reasons for this disparity, the reader may consult other articles in this volume. For purposes of this synthesis, a reanalysis of the normal curve equivalent (NCE) gains over time in each of the three programs was conducted, using the data reported in the Ramírez study. The reanalysis confirmed the patterns described in the final report.

After two years of schooling (K-1), students in all three programs (immersion strategy, early-exit, and late-exit) appeared to be doing relatively equally well at the end of the school year in first grade on the standardized Comprehensive Tests of Basic Skills (CTBS) English Reading and English Mathematics tests. They had

not reached national norms at the 50th NCE by any means, but the student results were comparable across the three programs when standard error of the means are taken into consideration. However, by the end of third grade, stronger differences between the programs began to emerge. Students in the immersion strategy programs, who received no L1 support, were losing in NCEs relative to the national norm group, and in some cases, were losing at a greater rate of loss with each succeeding year. Students in the early-exit programs, who received some L1 support in Grades K-2, were slightly gaining in NCEs relative to the norm on English reading and slightly losing on English math. In other words, they were keeping pace with the norm group but not, by third grade, catching up. Students in the late-exit programs, who received L1 support for Grades K-6, by third grade were keeping pace with the national norms in both English reading and math, and in third grade they were starting to make significant gains to *catch up with the norm group*. In the upper elementary school grades, the late-exit students made enough gains at a greater pace than the norm that they were able to make it to the 51st NCE in English math by sixth grade and to the 45th NCE in English Reading by sixth grade. Looking at patterns of *gains across time* as expressed in NCEs for the three programs, the immersion-strategy students were not likely ever to catch up to the norm; whereas the late-exit students demonstrated the ability to close the gap eventually between native speakers' performance and language-minority students' performance over time. "The growth patterns are surprisingly consistent across content areas.... Over and over again, those students who began their schooling with substantial amounts of instruction in their primary language and were exposed to the gradual introduction of English for instruction realized the greatest growth in skills" (Ramírez et al 1991, Vol. II p. 639).

Generalizations on long-term academic achievement in L2

How do the longitudinal findings in the Ramírez study compare to other long-term studies on language-minority students' academic achievement? Almost all the research looking at long-term results over at least a period of four years, including the Ramírez study, has found that *the greater the amount of L1 instructional support for language-minority students, combined with balanced L2 support, the higher they are able to achieve academically in L2 in each*

succeeding academic year, in comparison to matched groups being schooled monolingually in L2. In spite of the large number of studies that have been criticized for their methodological flaws (e.g. see methodological reviews in Baker & de Kanter 1983; Dulay & Burt 1979; McLaughlin 1985; Willig 1985; Zappert & Cruz 1977), there is strong research evidence from reasonably well conducted studies that balanced L1-L2 support over at least a six year period may make a substantial positive difference in language-minority student achievement, eventually allowing these students to catch up with the native-speaking norm group.

The key to examining this body of research is to examine students' patterns of gain over time, relative to the native-speaking norm group. Typically, language-minority students who are limited in English proficiency score at extremely low levels on standardized tests in English normed on native speakers. These tests are clearly not appropriate measures in the first couple of years of L2 development, and often school personnel exempt these students from norm-referenced tests because the tests will underestimate their true ability. Cross-sectional studies rarely control for number of years of students' exposure to L2. Thus most studies on language-minority students report extremely depressed standardized test scores when tested in L2 in the initial years of participation in a program.

Patterns of typical performance that can be seen in study after study are that in the first two or three years of a program, children schooled monolingually in L2 appear to make faster gains in L2 than comparable children being schooled bilingually. But typically around the third or fourth year (sometimes the fifth year), the bilingually schooled children begin to catch up in L2 to the monolingually schooled children. In other words, the bilingually schooled students start making greater gains in L2 than the monolingually schooled students. What happens after that is the crucial point. If a study is conducted for enough years (five or more), not only do the bilingually schooled students outperform their comparison group, making much greater gains, but they begin to reduce the gap between their performance and the norm group performance on standardized tests, achieving as high or higher than 50 percent of native speakers on a given test. In contrast, the most typical performance seen in scores of monolingually schooled language-minority students is that they appear to do well in the early grades, but their performance fails to match that of the norm group

and gains go down as they reach upper elementary and especially secondary schooling.

Guidelines for interpreting the studies. The studies to be presented in this article provide examples of some long-term research conducted in the United States over the past decade that compare bilingually schooled language-minority students to comparable monolingually schooled language-minority students. These studies also provide comparisons to national norms on standardized tests. While the latter is an inappropriate comparison in the first years of L2 exposure, the ultimate goal of U.S. school programs is for all students to have an equal opportunity to receive a comprehensive academic education at all grade levels. Since in the U.S., standardized tests are still used for decision-making, language-minority students want to be able to compete successfully in the long term with native speakers on whatever measures are used to determine access to educational programs. To ignore students' performance on these tests is to deny them access to a meaningful education and equal opportunity to benefit in life from their education.

For the sake of brevity, selected data from the studies are presented in this article in the form of tables. These tables cannot begin to capture the complexity of each study or the wealth of data that is reported in some of these studies, such as student background variables, instructional variations, actual classroom use of the two languages, administrative and community support, and all the other complexities of actual implementation. For this information, the reader should consult the original studies. Instead, this simplified picture provides a comparative glimpse across studies, with each table focused on one program variation in the use of L1 for instructional purposes. Some of the published summaries of the studies do not provide enough information to judge carefully methodological soundness. This review is intended to help educators know that at least these studies are being conducted and that there is a growing knowledge base in our field that eventually should give us a clearer picture of language-minority student academic achievement across time.

All the findings presented in the tables are based on national norms, reported in normal curve equivalents (NCEs). This provides a comparative look at the performance of language-minority students in each type of program to the national average for native speakers'

performance on these standardized tests. NCEs are reported rather than percentiles, because percentiles are not equal-interval data. When comparing student performance or performing statistical operations (such as simple addition and subtraction), only equal-interval scales can be used. For this reason, all studies that reported grade-equivalent scores were eliminated from this review, as the grade-equivalent scores used in these studies were not equal-interval data and are useless for comparisons.

The test scores reported in the tables are limited to English reading and English mathematics, as these are the most commonly reported scores across all studies, providing one measure of students' L2 language development and one measure of their L2 content development. Many of these studies report findings on a standardized English language arts test as well as English reading. However, reading scores are chosen here as a more valid predictor of L2 thinking skills than language arts scores. Most language arts tests measure only discrete points of language, such as spelling, punctuation, and simple grammar. A reading test typically measures cognitively more complex language skills through measures of reading comprehension and vocabulary analysis, and a reading test is considered to be a better predictor of students' L2 academic performance at secondary and postsecondary levels of study. Most of these studies also provide results from tests measuring L1 language and cognitive development that are not included in this brief review, although in programs with L 1 support these results are generally very impressive in demonstrating students' strong academic achievement in L1. Also not included in the tables, for lack of space, are the results of the matched comparison groups performance on the tests. In every study, the comparison group being schooled monolingually scored lower on the tests, and in most findings, the difference was statistically significant.

In order to be consistent, every long-term study from the 1980s that could be located has been included in this review, but some of the studies have too small a number of students for the results to be generalizable. Studies with very small student samples are labeled as such in the descriptive comments following each table. Also, all of these studies are cross-sectional; thus it is likely that these are conservative estimates of language-minority students' actual performance. For purposes of these comparisons, it must be assumed that students who were tested received the program treatment from kindergarten or first grade until the year of testing.

But in fact, most studies include some newly arriving students who are just beginning their L2 development in second, third, fourth, or fifth grade, mixed in with the test score results of all those who received the program treatment for all of their schooling. This means that in a carefully controlled longitudinal study, such as the Ramírez study, actual language-minority student performance would probably be somewhat higher than findings in the cross-sectional studies.

The tables in this article are presented in an order according to the amount of L1 support provided in each program type, from those programs that provide the strongest L1 support to those that provide the least L1 support for language-minority students. Each table includes information on the location where the study was conducted, the author and date of a publicly-available reference on the study, the language-minority group being given the special program treatment and being tested, the number of years of L1 support and the percentage of L1 support for each year, the highest grade at which students were tested in the study, the name of the test, and the test results at this grade level in English reading and English mathematics, reported in NCEs.

Two-way bilingual education. Table 1 presents long-term findings from four two-way bilingual programs. No two-way bilingual schools were included in the Ramírez study. Two-way bilingual education is an integrated program in which language majority students work together academically with language-minority students, learning language and content through both L1 and L2. Lindholm (1987) has described this program model in some detail and a current survey of two-way bilingual schools in the U.S. has just been completed (Christian 1992). Federal funds are available for this type of program within the category of developmental bilingual education.

For decisions on the proportion of L1 and L2 instruction, some U.S. two-way bilingual programs use the early total immersion model of Canada, beginning with all instruction in the *minority* language in kindergarten and gradually adding instruction in English each year until by fourth grade, there is half a day of academic instruction in each language; while other U.S. programs implement half a day of academic instruction in each language from the first year of schooling, labeled "partial-immersion" in Canadian bilingual education. As a parenthetical comment here, it is important to

understand that *all immersion programs in Canada are bilingual programs* with full support for two languages for all grade levels, K-12.

The two-way bilingual program model has strong potential for high academic achievement of all students by lessening social distance and unequal social status relations between majority and minority language students. No exit is needed from this type of program, since a two-way bilingual class is a mainstream class, integrated with language majority speakers. Unfortunately, to date, very little research has been conducted on students' long -term academic achievement in this type of program, but a number of studies are in progress (Christian 1992). The few studies that have been conducted over a period of at least five years, such as the data presented in Table 1, have found that this may be a very effective program model for all students (Collier 1989a, 1992; Crawford 1989; Lindholm 1987).

Table 1
Outcomes of two-way bilingual programs

Location and Reference	LM Student Group	L1 Support	Grade Tested, Test	Test Results
Arlington, VA (Rhodes &Barfield, 1991)	Hispanic (50% of class)	K-650%	Grade 4 ITBS	Eng. Reading 52nd NCE Eng. Math 69th NCE
Washington DC (Brief report in Crawford, 1989)	Hispanic (50% of class)	K-6 50%	Grade 6 CTBS	Eng. Reading 77th NCE Eng. Math 85th NCE
Chicago, IL (Brief report in Lindholm, 1987)	Hispanic (60% of class)	K-8 50%	Grade 8 ITBS	Eng. Reading above 50th NCE
San Diego, CA (Herbert, 1987)	Hispanic (60% of class)	K-3 80% 4-6 50%	Grade 6 CTBS	Eng. Reading 51st NCE Eng. Math 64th NCE

As can be seen in Table 1, students in a two-way program for at least 4-5 years tend to score very high on standardized tests in English. In two of these schools, language-minority students typically reach the 50th NCE when tested in L2 by third grade and

continue to score still higher in each succeeding grade. However, two caveats in reading this table are important. First, the test results for two of the programs above (Arlington, VA, and Washington, DC) represent the whole class performance; no breakdown by language group was available, but research is in progress and reports will be available in 1993. Second, each of these studies represents too small a number of students (around 20-30 students) for the results to be generalizable. But from these initial long-term results, one can conclude that the results for two-way bilingual programs look promising.

Late-exit bilingual education. Table 2 presents some long-term findings from five late-exit bilingual programs designed for language-minority students. For purposes of this article, the term "late-exit bilingual program" will be used, as adopted in the Ramírez study, to refer to those bilingual programs that provide considerable support for language-minority students' L1 throughout the elementary school years. In the U.S., these programs have more traditionally been referred to as "maintenance" models, although a strict definition of maintenance would imply L1 support for grades K-12; for example, all immersion programs in Canada are maintenance bilingual education programs.

Similar to two-way programs, except for the lack of majority language students attending the classes, a late-exit (or maintenance) program might choose to adopt a balanced 50-50 L1-L2 ratio of instructional language from kindergarten through Grade 6. This model of language use typically introduces L1 and L2 reading in the same year. A second model separates the introduction of literacy in each language, with primary language reading taught first and L2 literacy introduced in second or third grade. The California State Department of Education chose in the early 1980s to adopt the second type of late-exit program as their preferred model of instruction, after following the research results of several "case study" schools that changed instructional practices to follow the theoretical principles set forth by the state office. In addition to introducing primary language literacy before teaching second language literacy, California also emphasizes the importance of separating the two languages of instruction, with no use of translation, and teaching language through the content areas (California State Department of Education 1982). Four of the five studies presented in Table 2 provide results from late-exit programs

in California; many other studies are in progress in other states, but none have yet published long-term results.

Table 2
Outcomes of late-exit (or maintenance) bilingual programs

Location and Reference	LM Student Group	L1 Support	Grade Tested, Test	Test Results
Baldwin Park, CA (Krashen & Biber, 1988)	Hispanic N=39	K-2 80% 3-6 50%	Grade 5 CTBS Grade 8 CTBS	Eng. Reading 51st NCE Eng. Math 59th NCE Eng. Reading 44th NCE Eng. Math 57th NCE
Calexico, CA (Krashen & Biber, 1988	Hispanic	K-2 80% 3-4 50% 5-6 20%	Grade 6 CTBS	Eng. Reading 44th NCE Eng. Math 55th NCE
Los Angeles, CA (Krashen & Biber, 1988)	Hispanic	K-2 80% 3-6 50%	Grade 6 CTBS	Eng. Reading 50th NCE Eng. Math 59th NCE
San Francisco,CA (Krashen & Biber, 1988)	Hispanic N=12	K-1 80% 2-4 50% 5-6 25%	Grade 3 CTBS Grade 6 CTBS	Eng. Reading 51st NCE Eng. Math 52nd NCE Eng. Reading 52nd NCE Eng. Math 55th NCE
Arizona (Medina, Saldate & Mishra, 1985)	Mex-Amer	K-5	Grade 12	Eng. Reading & Eng Math above 50th NCE

The data for the late-exit bilingual programs are very encouraging, especially when compared to these schools' earlier academic achievement before bilingual services of this type were initiated. There are, again, some caveats in interpreting the data in Table 2. First, the test results in two of the studies (Calexico, CA, and Los Angeles, CA) represent test scores for the whole school and thus cannot be assumed to be a direct result of program treatment. However, these two schools were among the lowest-scoring schools in the state before the new bilingual program, and 60-66% of the student body participated in the new program. Second, the data from San Francisco cannot be considered generalizable, because the number of students tested (12) was insufficient. The study by Medina, Saldate and Mishra (1985) examining twelfth grade test results from graduates of a K-5 bilingual program

provides a rare glimpse of secondary language-minority students' performance. More studies following students' academic achievement at secondary level are badly needed to see if the dramatic gains from late-exit programs can be sustained.

Early-exit bilingual education. Table 3 presents findings from five early-exit (transitional) bilingual programs designed for language-minority students. This program model is the most common model for L1 support in the U.S., with limited L1 instruction in basic literacy and cognitive development typically provided over a 2-4 year period. Because the largest amounts of federal funding for direct services to students have been available for this program model, hundreds of program evaluations have been conducted. However, rarely are these studies made public, so that, once again, there is not an extensive research base on language-minority student achievement in early-exit programs. Especially missing are studies on language-minority student academic achievement after exiting from these programs; results of student performance at the end of second or third grade do not provide a sufficiently long-term picture to predict students' success in later grades. Three of the five studies presented in Table 3 examined test results after exit from the program, in fifth or sixth grade.

Again, caveats are needed in the interpretation of the data presented in Table 3. The Tucson, AZ, study has too small a number of students (25) for the results to be generalizable. However, this is one of a series of studies being conducted by a growing number of researchers in the Southwest (Medina, de la Garza, Saldate, Mishra, Medrano, Escamilla, Lindholm, and others) who are working on longitudinal and cross-sectional studies examining language-minority students in early-exit, late-exit, and two-way bilingual programs.

The El Paso, TX, studies presented in Table 3 are interesting in their examination of two early-exit programs that differed somewhat in the amount of L1 support provided for students, but differed greatly in the instructional style used in classrooms. One program was taught rather traditionally, using sequenced phonics instruction for teaching L1 and L2 literacy and an audiolingual grammar-based approach for teaching L2. The second program provided less L1 support, but L1-L2 literacy was taught using more current whole language approaches and L2 was taught using the Natural Approach in interactive, content-based classes. While the standardized test

scores by fifth grade do not indicate the strong difference in student achievement between the two programs, other performance-based and state criterion-referenced tests found significant differences in favor of the more interactive/experiential classes. Teachers in the more traditional program have now been trained in the newer approaches and are gradually changing their teaching style.

Table 3
Outcomes of early-exit (or transitional) bilingual programs

Location and Reference	LM Student Group	L1 Support	Grade Tested, Test	Test Results
El Paso, TX (Office for Research &Evaluation, 1989)	Mex-Amer	L1: K-1 80% 2-3 40-60% L2: Phonics & ALM	Grade 5 ITBS	Eng. Reading 30th NCE Eng. Math 43rd NCE
El Paso, TX (Office for Research &Evaluation, 1989)	Mex-Amer	L1: K-3 20-30% L2: Whole language &NA	Grade 5 ITBS	Eng. Reading 34th NCE Eng. Math 44th NCE
Tucson, AZ (Medina & de la Garza, 1989)	Mex Amer N=25	L1: Grade: 1: 75% 2: 70% 3: 50%	Grade 2 CAT Grade 3 CAT	Eng. Reading 58th NCE Eng. Reading 55th NCE
Yakima, WA (McConnell & Kendall, 1987)	Hispanic	L1: 3 yrs.	Grade 6 WRAT	Eng. Reading 50th NCE Eng. Math above 50th NCE
Pullman, WA (McConnell, 1981)	Hispanic migrants	L1: K-3 + extra.	Grade 3 WRAT	Eng. Reading 50th NCE Eng. Math 61st NCE

Two other studies of early-exit programs not presented in Table 3 are more difficult to present in the table because they used statistical procedures to examine *NCE gains in pre-post testing* of students each year over several years. This is similar to what is typically done in Chapter I and current Title VII evaluations. In one study, Egan and Goldsmith (1981) examined patterns over several years of student data from transitional bilingual programs for

Mexican-American students from 38 school districts in the state of Colorado. These programs provided some L1 support for K-3 and were state mandated. Comparing program data across school districts, they found that in 34% of the bilingual programs, students had averaged a gain of 7 or more NCEs, 64% of the programs averaged a student gain of at least 3 NCEs, and 87% of the programs had zero NCE gains or greater. Since a zero NCE gain means that students are keeping up with the pace of native speaker gains, a gain of 3 NCEs indicates that language-minority students are outpacing the norm, and a gain of 7 NCEs indicates a very significant gain (one-third of a national standard deviation). They concluded that theirs was a success story, at least for the grades examined. A next step, though, is to continue following students' progress in upper elementary and secondary grades.

A second study by Baecher (1989) examined pre-post NCE gains of recently arrived immigrants from Central and South America, some with very limited past schooling, who received strong L1 support for three years in a program in Port Chester, NY. Although this program was labeled two-way, these students only attended classes with native English speakers twice a week; thus for this discussion, the program is more appropriately classified as early-exit. Baecher reported very statistically significant achievement gains among these immigrants on the Stanford Diagnostic Reading Test in English, with gains of 13 NCEs in Grade 3, 16 NCEs in Grade 4, and 8 NCEs in Grade 5. However, because newly arrived students were included at each grade level for each year of the program and the total number of students receiving the program treatment was so small, the group in Grade 5 had only reached the 16th NCE on this standardized test in L2. The same students scored above the 50th NCE on standardized tests in Spanish. The number of students in this study (13-19) is too small to be generalizable, and there is substantial uncertainty in each mean because of small sample size. But this study demonstrates the difficulty of both controlling the variable of amount of exposure to L2 and still having a large enough sample for results to be generalizable. This study also shows how even with very dramatic gains, it takes quite a few years for limited-English-proficient students to reach native speaker norms on standardized L2 tests.

Structured immersion and ESL. Table 4, the last table of long-term studies from the 1980s presented in this synthesis,

Table 4
Outcomes of programs with no L1 support

Location and Reference	LM Student Group	Type of Program	Grade Tested, Test, Results
Uvalde, TX (Becker & Gersten, 1982; Gersten & Woodward, 1985)	Mex-Amer	Structured Immersion	Grade 3, Eng. Reading 39th NCE MAT Eng. Math 47th NCE Grade 5, Eng. Reading 29th NCE MAT Eng. Math 35th NCE Grade 6, Eng. Reading 29th NCE MAT Eng. Math 32nd NCE
Fairfax Co., VA (Collier & Thomas, 1989)	Immigrants: 65% Asian 20% Hispanic 75 languages performing on grade level in L1 when arrived in U.S.	ESL	SRA English Reading: After 5 yrs of L2 After 6 yrs Grade 4: 51st NCE Grade 6: 51st NCE 51ST NCE Grade 8: 46th NCE 47th NCE Grade 11: 25th NCE 31st NCE SRA English Math: After 2 yrs of L2 After 6 yrs Grade 4: 64th NCE Grade 6: 66th NCE 68th NCE Grade 8: 73rd NCE 73rd NCE Grade 11: 53rd NCE 59th NCE
Toronto, Canada (Cummins, 1981a)	Immigrants: many languages	ESL	English Vocabulary: Tested in Grades 5, 7, 9 After 5 yrs of L2 schooling: Reaching 40th-45th NCE After 7-9 yrs of L2 schooling: Reaching 45th-50th NCE

combines two types of programs with no L1 support. Structured immersion (labeled "structured immersion strategy" in the Ramírez study) is a program developed in the early 1980s in the U.S. that provides all instruction in L2. It does not in any way resemble immersion programs in Canada. In the first year of the program,

students are allowed to speak their L1 in the classroom and the teacher has sufficient knowledge of students' L1 to understand them, but the teacher speaks only English. In English-as-a-SecOnd language (ESL) programs, all instruction is in English, and usually a variety of primary languages are represented among students in the class. There are very few structured immersion programs in the U.S., but there are thousands of ESL programs, as this is the most common special program provided for limited~English-prOficient students of many language backgrounds. Both structured immersion and ESL provide short-term support in L2 for 2-3 years from teachers who have been specially trained in L2 acquisition and instructional methods for teaching ESL. In recent years, the field has changed radically, from teaching methods that provided a structured, sequenced curriculum in a teacher-centered class to experiential methods that teach language through content in a student-centered, interactive, problem-solving class.

The three studies presented in Table 4 are the only published studies examining language-minority student academic achievement over a period of several years after receiving ESL or structured immersion support. More studies are clearly needed. It is important to note that these studies were conducted before the change in ESL teaching methods took place. One study is included from Canada, because so few studies are available.

Examining the data presented in Table 4, students from the structured immersion program in Uvalde, TX, were losing ground, relative to the norm group, with each succeeding upper elementary grade level. In addition to the Ramírez study, only one other study on structured immersion has been published (Gersten 1985), but this study is not included in the table because the CTBS tests used were not on-grade-level tests and the number of students tested (16) was too small to be generalizable.

The Fairfax County, VA, study analyzed records of 2,014 immigrants, and the Toronto, Canada, study included 1,200 immigrants. Both of these studies controlled the variable of number of years of L2 exposure and found that students of ages eight and above, studying all in L2 with no L1 support at school, took a minimum of 5-9 years to reach the 50th NCE on standardized English reading and vocabulary tests. Collier and Thomas (1989) also controlled the variable of L1 schooling in home country and found that students arriving in the U.S. at a very young age (who had received little or no L1 schooling in their home country) took 7-

10 years to reach the 50th NCE in English reading, projecting their rate of gain from the six years of the study.

The other age group in the Collier and Thomas (1989) study in need of more support were high school students. The eighth graders (who had had at least three years of schooling in their home countries before coming to the U.S.) were doing relatively well in English reading (at the 45th NCE) and extremely well in math (at the 73rd NCE) after only two years of English exposure. But these well-schooled immigrant eleventh graders only made it to the 31st NCE in English reading, the 37th NCE in science, and the 38th NCE in social studies after six years of all-English schooling in the U.S. Almost no long-term research has been conducted on language-minority students in the U.S. at secondary level, to follow their academic growth across time; much more research is needed for this age group, analyzing their progress in the mainstream after receiving various types of special program support.

Summary

What do all these findings tell us? The general patterns of these studies from the 1980s on language-minority students' academic achievement in U.S. schools confirm the generalization stated earlier in this article that *the greater the amount of L1 instructional support for language-minority students, combined with balanced L2 support, the higher they are able to achieve academically in L2 in each succeeding academic year, in comparison to matched groups being schooled monolingually in L2*. The Ramírez study found this pattern, although with only four years of longitudinal data collected, the findings can only point to the trend seen in the data and are therefore stated somewhat tentatively. Data from two-way and late-exit bilingual programs look the most promising. Early-exit bilingual programs provide important support, but may not be as successful in long-term student academic achievement as late-exit and two-way bilingual programs. In programs with no L1 support, it may take a very long time for language-minority students to reach national norms in L2.

The studies presented in this article are not the only ones that examine long-term data. This review was limited to all published studies conducted over the past decade on language-minority students in the U.S. Nineteen other long-term studies from the 1970s and early 1980s (reported in Cummins 1981b; Fulton-Scott & Calvin 1983; Tempes et al, 1984; and Troike 1978) found very

similar patterns but are not included here because most data from these studies was reported in grade equivalent scores, which are not equal interval scales and are therefore inappropriate for comparisons. Nevertheless, it is interesting to note that these 19 studies found the same type of pattern in student performance--that students taught through a minority language for part of the school day may experience an initial lag in L2 skills which usually disappears by the middle grades of elementary school and in many cases the bilingually schooled students demonstrate superior performance by the end of elementary school, in comparison to matched monolingually schooled students.

Another group of studies examining language-minority students' academic achievement across several years was examined by Willig (1985), who conducted a meta-analysis of 23 studies in the U.S. from the 1970s. These studies examined the effectiveness of transitional bilingual education compared to all-English instruction for language-minority students. Willig's carefully conducted statistical reanalyses of the data of these studies revealed a significant difference in favor of providing L1 support for language-minority students. In her conclusions, she strongly recommends that future studies should be more carefully designed to avoid some of the methodological mistakes of previous research.

Still another group of studies not reported in this article that have found the same pattern of school performance among language-minority students schooled bilingually are studies conducted m other countries. The interested reader can find syntheses of some of these studies in Collier (1989b), Cummins (1981b, 1983), Dutcher (1982), McLaughlin (1985), and Skutnabb-Kangas (1981). Another group of studies with interesting parallels are the analyses of language majority students being schooled bilingually in immersion programs in Canada. These studies have consistently found that language majority students taught through a minority language for part of the school day may experience an initial lag in L1 skills in the first 2-3 years of the program which usually disappears by the middle grades of elementary school, and in many cases the bilingually schooled students demonstrate superior academic performance by the end of elementary school, in comparison to monolingually schooled students (e.g. see syntheses in California State Department of Education 1984; Collier 1992; Cummins & Swain 1986; Genesee 1987; Harley, Allen, Cummins & Swain 1990; Swain & Lapkin 1981).

One more group of studies that provide additional evidence for the importance of some L1 support for language-minority students are studies that have examined patterns of school achievement among immigrants who received some schooling in their home country before emigrating to the host country. These studies have found that immigrants who received at least 2-3 years of home language instruction before emigration did better in L2 schooling than those immigrants who received all their schooling in L2 (e.g. see Collier & Thomas 1989; González 1985; Skutnabb-Kangas 1979; Troike 1986).

In conclusion, it is important to remember that this synthesis of studies has focused on only one of many possible program, instructional, and student background variables that may affect student academic achievement. Future research will likely find that it is not L1 support alone that may influence student achievement but L1 support in combination with many other factors (e.g. see Dolson 1985; Valdez Pierce 1991). We still have many unanswered questions and need additional research from methodologically well conducted studies. The Ramírez study provides some path-breaking analyses, but more years of student achievement data are needed to have clearer answers to the effectiveness questions asked. However, this study does add more research evidence to the growing knowledge base in language-minority education that L1 support may be one of the key variables in long-term language-minority student success.

References

Baecher, R.E. 1989, March. *Links between bilingualism, achievement, and psychosocial classroom environment among bilingual and monolingual students.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Baker, K.A., & A. de Kanter. Eds. 1983. *Bilingual education: A reappraisal of federal policy.* Lexington, MA: D.C. Heath.

Becker, W.C., & R. Gersten. 1982. A follow-up of Follow Through: The later effects of the direct instruction model on children in fifth and sixth grades. *American Educational Research Journal, 19,* 75-92.

Borg, W.R., & M.D. Gall. 1982. *Educational research* (4th ed.). New York: Longman.

California State Department of Education. 1982. *Basic principles for the education of language minority students: An overview.* Sacramento, CA: Author.

California State Department of Education. 1984. *Studies on immersion education: A collection for United States educators.* Sacramento, CA: Author.

Christian, D. 1992. *Two-way bilingual programs in the United States, 1991-1992.* Washington, DC: Center for Applied Linguistics.

Collier, V.P. 1989a, March. *Academic achievement, attitudes, and occupations among graduates of two-way bilingual classes.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

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

Collier, V.P., & W.P. Thomas. 1989. How quickly can immigrants become proficient in school English? *Journal of Educational Issues of Language Minority Students, 5*, 26-38.

Crawford, J. 1989. *Bilingual education. History, politics, theory, and practice.* Trenton, NJ: Crane Publishing.

Cummins, J. 1981a. Age on arrival and immigrant second language learning in Canada: A reassessment. *Applied Linguistics, 11,* 132-149.

Cummins, J. 198 lb. The role of primary language development in promoting educational success for language minority students. In *Schooling and language minority students: A theoretical framework* (pp. 3-49). Sacramento, CA: California State Department of Education.

Cummins, J. 1983. *Heritage language education. A literature review.* Toronto: Ministry of Education, Ontario.

Cummins, J., & M. Swain. 1986. *Bilingualism in education*. New York: Longman.

Dolson, D. 1985. Bilingualism and scholastic performance: The literature revisited. *NABE Journal, 10* (1), 1-35.

Dulay, H., & M. Burt. 1979. Bilingual education: A close look at its effects. *Focus*, No. 1. Washington, DC: National Clearinghouse for Bilingual Education.

Dutcher, N. 1982. The use of first and second languages in primary education: Selected case studies. World Bank Staff *Working Paper,* No. 504. Washington, DC: The World Bank.

Egan, L.A., & R. Goldsmith. 1981. Bilingual bicultural education: The Colorado success story. *NABE News, 4* (3), 1, 4, 12-14.

Fulton-Scott, M.J., & A.D. Calvin. 1983. Bilingual multicultural education vs. integrated and non-integrated ESL instruction. *NABE Journal, 7* (3), 1-12.

Genesee, F. 1987. *Learning through two languages: Studies of immersion and bilingual education.* Cambridge, MA: Newbury House.

Gersten, R. 1985. Structured immersion for language minority students: Results of a longitudinal evaluation. *Educational Evaluation and Policy Analysis, 7,* 187-196.

Gersten, R., & J. Woodward. 1985. A case for structured immersion. *Educational Leadership, 43* (1), 75-79.

González, L.A. 1985. *The effects of first language education on the second language and academic achievement of Mexican immigrant elementary school children in the United States.* Doctoral dissertation, University of illinois at Urbana-Champaign.

Harley, B., Allen, P., Cummins, J., & M. Swain. (Eds.). 1990. *The development of second language proficiency.* Cambridge: Cambridge University Press.

Herbert, C. 1987. *San Diego Title VII Two-way Bilingual Program.* San Diego, CA: San Diego Unified School District.

Krashen, S., & D Biber. 1988. *On course: Bilingual education's success in California.* Sacramento, CA: California Association for Bilingual Education.

Lindholm, K.J. 1987. *Directory of bilingual immersion programs: Two-way bilingual education for language minority and majority students*. Los Angeles: Center for Language Education and Research, University of California, Los Angeles.

McConnell, B.B. 1981. *Long-term effects of bilingual instruction.* Pullman, WA: Bilingual Mini Schools.

McConnell, B.B., & J.R. Kendall. 1987, April. *Application of the cohort model to evaluate bilingual programs: The "BELEPS" program.* Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

McLaughlin, B. 1985. *Second language acquisition in childhood: Vol. 2. School-age children* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Medina, M., & J.V. de la Garza. 1989. Initial language proficiency and bilingual reading achievement in a transitional bilingual educational program. *NABE Journal, 13* (2), 113-125.

Medina, M., Saldate, M., & S. Mishra. 1985. The sustaining effects of bilingual instruction: A follow-up study. *Journal of Instructional Psychology, 12* (3), 132-139.

Office for Research and Evaluation, El Paso, TX. 1989. *Bilingual education evaluation: The fifth year in a longitudinal study.* El Paso, TX: El Paso Independent School District.

Ramírez J., S. Yuen, D. Ramey & D. Pasta. 1991. *Final Report: Longitudinal study of structured English immersion strategy, early-exit and late-exit bilingual education programs for language-minority children.* (Vol. I) (Prepared for U.S. Department of Education). San Mateo, CA: Aguirre International. No. 300-87-0156.

Ramírez, J., Pasta, D., Yuen, S., Ramey, D. & D. Billings. 1991. *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit bilingual education programs for language-minority children.* (Vol. II) (Prepared for U.S. Department of Education). San Mateo, CA: Aguirre International. No. 300-87-0156.

Rhodes, N.C., & S.C. Barfield. 1991. *Review of the fifth year of the partial immersion program at Key Elementary School, Arlington, Virginia.* Washington, DC: Center for Applied Linguistics.

Skutnabb-Kangas, T. 1979. *Language in the process of cultural assimilation and structural incorporation of linguistic minorities.* Washington, DC: National Clearinghouse for Bilingual Education.

Skutnabb-Kangas, T. 1981. *Bilingualism or not: The education of minorities.* Clevedon, England: Multilingual Matters.

Swain, M., & S. Lapkin. 1981. Bilingual education in Ontario.' A decade of research. Toronto: Ontario Institute for Studies in Education.

Tempes, F., Burnham, L., Piña, M., Campos, J., Matthews, S., Lear, E., & C. Herbert. 1984, January. *Implementing theoretically sound programs: Do they really work?* Paper presented at the annual conference of the California Association for Bilingual Education, San Francisco, CA.

Troike, R.C. 1978. Research evidence for the effectiveness of bilingual education. *NABE Journal, 3* (1), 13-24.

Troike, R.C. 1986. *Improving conditions for success in bilingual education programs.* Paper presented to the Committee on Education and Labor, U.S. House of Representatives.

Valdez Pierce, L. 1991. *Effective schools for language-minority students.* Washington, DC: Mid-Atlantic Equity Center, American University.

Willig, A. 1985. A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55,* 269-317.

Zappert, L.T., & B.R. Cruz. 1977. *Bilingual education: An appraisal of empirical research.* Berkeley, CA: Bahía.