

**Nothing Matters?:
A Critique of the Ramírez, et al.
Longitudinal Study of Instructional Programs
for Language-Minority Children**

Christine H. Rossell
*Political Science Department
Boston University*

Abstract

The Ramírez, Pasta, Yuen, Ramey & Billings 1991 study analyzing the achievement of 1,054 language-minority children in structured immersion, early-exit and late-exit bilingual programs has several serious flaws which lead me to conclude that we cannot place any confidence in the finding of no consistent difference in the achievement of children regardless of how much Spanish or English is used in instruction.

Introduction

The Ramírez, Pasta, Yuen, Ramey & Billings (Ramírez, et al. Vol. II, 1991) *Longitudinal Study of Structured English Immersion Strategy, Early-Exit and Late-Exit Transitional Bilingual Education Programs for Language-Minority Children* of approximately 1,054 children¹ in 9 school districts, 46 schools, and 136 classrooms across 5 grades for immersion/early-exit and 7 grades for late-exit (Vol. I, p. 54) over four years, was intended to overcome the many defects of bilingual education research that analysts have been lamenting for decades—especially the lack of a control group and the lack of statistical control for pre-treatment differences in most studies. Although the Ramírez et al. study cost \$4.5 million, it does not answer many of the most important questions asked in the Request For Proposal and, therefore, does not deliver what it should for that price. Nevertheless, this is by far the most impressive study of bilingual education since the American Institute of Research Analysis (Danoff, Arias, Coles & Everett 1977; Danoff, Coles,

¹ The number of students in this study varies with the analysis conducted. The total number of students with pretest data in all programs is only 486 for the most important analysis - that of achievement from grades 1 to 3 (Ramírez, et al. 1991, Vol. II, pp. 60-62).

McLaughlin & Reynolds 1978). The report is 1,148 pages and consists of two volumes. Volume I describes the characteristics of the programs, administrators, teachers, students and their families. Volume II presents the statistical analysis of achievement outcomes for three whole-class, all day programs for limited-English-proficient children.

The three programs that are compared are 1) a structured immersion program with all-English instruction and a bilingual teacher, 2) an early-exit bilingual program with a bilingual teacher but a goal of transitioning children to English as early as possible, and 3) a late-exit bilingual program whose goal is to develop students' Spanish as well as English proficiency.

There are four major problems with this study that limit its policy usefulness. The biggest shortcoming is that Ramírez, et al. (1991) never compare the achievement of children in the late-exit bilingual education program—the one with the most Spanish instruction—to that of children in the immersion and early-exit bilingual education programs.² The second major deficiency of this study is that they did not collect achievement data for children in the immersion and early-exit programs for grades 5 and 6 because they expected these children to have been exited from their programs by then and because the immersion programs were so new that there were little or no 5th and 6th grade data. The third major problem is that they looked at nominal program types—immersion~ early-exit, late-exit—rather than defining the program by its most important characteristic—the extent of English used in each subject. The fourth major problem is that they did not study at all the most widely implemented, and probably least expensive, strategy for teaching LEP children—regular classroom enrollment with English as a

² This is because they were unable to locate immersion and late-exit programs in the same districts. They unconvincingly argue that they could not compare different programs in different districts because they would not be able to distinguish differences between programs from differences between school districts. This argument is unconvincing because they do compare immersion and early-exit programs in different schools and as they found out, there is more variation between schools than between districts (see Ramírez et al. Vol. II, p. 45). In addition, they compare late-exit programs in different districts although there were large difference in the characteristics of the programs.

second language (ESL) pull-out instruction.³ Thus, we will never know how the most widely-used technique in the U.S. compares to immersion, a technique hardly implemented at all, and to early-exit and late-exit bilingual programs. There are additional problems as well that are discussed in more detail below.

The Programs and students

Program model problem

While much of the study design is quite impressive, there were many decisions made that I believe were unfortunate. The authors, for example, unnecessarily limited the number of potential school districts with “immersion strategy” programs by requiring the teacher to be bilingual in the native tongue and English⁴ and that project classrooms have limited English proficient (LEP) students who all speak the same native tongue. Requiring these program elements not only eliminated most of the structured immersion programs in the U.S., but limited the generalizability and policy relevance of this study since this model would be impossible to implement widely in the U.S. with its diverse immigrant population. In addition, by eliminating most of the programs in the U.S. because they required these characteristics the researchers could not a) randomly select from among the existing immersion programs or b) obtain enough immersion programs to compare them to late-exit bilingual programs in the same district. One wonders how the five school districts with immersion programs that they did study differ from the many other school districts that were excluded.

³ This model differs from structured immersion in that with structured immersion students are in a special classroom with a specially trained teacher all day long. In ESL, they are in a regular classroom with a regular teacher, but pulled out for small class remediation with a specially trained teacher for about an hour a day.

⁴ The proponents of bilingual education in the U.S. make much of the fact that the teachers in the so-called Canadian “immersion” programs are bilingual. However, I would argue that since they rarely, if ever, use their second language ability in the most successful portion of the program - the kindergarten and first grade Immersion program - that their second language ability is a luxury, not a necessity, for the immersion program. (It may, however, be useful in communicating to Parents.) When the program becomes bilingual, of course, it is a necessity that the teachers be bilingual.

The researchers were also able to locate only five districts in the country that met the criteria for a developmental Spanish (i.e. late-exit) program; of these, only three were selected. This is almost certainly wrong⁵ because I can personally name more than six. The problem the researchers had is two-fold. First, they relied on the state education authorities and local school districts to accurately depict their programs. Second, many of the early-exit programs in the U.S. have as their official goal the transitioning of students into the native language, but in reality this does not occur. Thus, many early-exit programs are actually late-exit programs.

Ramírez, et al. further limited the number of late-exit programs they could analyze by requiring that the use of Spanish and English be differentiated by teaching staff, with teacher A only using Spanish and teacher B only using English. This criterion does not make sense to me since there are many valid late-exit programs that use one teacher all day long.

There are other problems with their criteria. They define an early-exit program as one in which there is no more than one hour a day of instruction in Spanish (Vol. I, p. 39). However, not only does the amount of Spanish instruction vary considerably from program to program, but the amount of Spanish used in an early-exit transitional bilingual program differs from the beginning of the program when it may be more than 50 percent to the end of the program when it may be 0. Which part of this "transition" are the authors referring to? And which part of this "transition" were the educational authorities thinking of when they responded? One can only wonder how many early-exit transitional bilingual education (TBE) programs were excluded from or included in this study because the state or school district authorities were thinking of one end of this continuum not the other.

Ramírez, et al. have similar problems with their definition of late-exit bilingual programs. They define a late-exit program as one with instruction in the Spanish language at least 50 percent of the day.⁶ But, as indicated above, there are many early-exit programs which in the beginning use Spanish more than 50 percent of the

⁵ One cannot say for sure this is wrong since the researchers do not give us the names of the districts they studied.

⁶ This study flipflops back and forth on whether the criterion is 40 or 50 percent instruction in the native tongue in late-exit programs.

time. Moreover, there is also a lot of variation not only from program to program, but across grades. Which part of this continuum were the authors thinking of? What were the respondents thinking of when they answered?

In fact, in Volume II they analyze the English language usage standard deviations for the three program models and find that only the immersion strategy teachers faithfully follow their model. Several of the early-exit programs more closely resembled the immersion strategy programs and a couple of late-exit programs resembled the early-exit model. The final sample of programs is thus neither representative of those in the U.S. nor clearly defined.

Program and student characteristics

Oddly enough, having required that the teachers in every program, including the immersion strategy, be bilingual in order to be selected, they did not require that they be either trained or certified in second language acquisition techniques. The result is that the late-exit teachers are better educated, more fluent in Spanish, and have more specialized training to work with language-minority students than teachers in immersion strategy or early-exit programs. This is a potential source of bias in the study since these variables are never entered into any equations predicting achievement.

The study is a combination of a cohort analysis in which different students are analyzed in different grades and a longitudinal analysis in which the same students are followed over time (three or four years depending on the cohort). Different students in different cohorts in the same program are treated as one child. Thus, in Volume I or II whenever K-6 data are presented for late-exit students and K-4 data are presented for early-exit/immersion students, they are not the same children within each program, something the reader is not kept adequately informed of.

Not only were the districts not randomly selected, but neither were the schools or teachers. In most cases, they “were selected arbitrarily by the district’s project director for participation in the study” (Ramírez, et al. 1991, Vol. I, p. 48). Thus, the final sample of schools and teachers may be biased in a way that not only limits external validity (generalizability), but internal validity (the accurate determination of causality). Indeed, in Volume II, Ramírez, et al. note that “There is evidence that some districts deliberately implemented immersion strategy programs in their ‘best’ or ‘worst’

schools”⁷ (Ramírez, et al. 1991, Vol. II, p. 84). The best that can be said about this sample is that it is one of convenience.

The data collected is quite comprehensive and impressive, and includes almost every known important variable. There is a problem, however, with the achievement data. Students were tested with one type of test in kindergarten but another type, California Test of Basic Skills (CTBS) in Spanish and English in the fall and spring of subsequent years. Thus there is no continuity between the pretest in kindergarten and the tests used in later grades. Because of this, the estimates of growth in achievement begin in the spring of first grade (almost two years after the start of the program) rather than in kindergarten.

The study’s great strength is that, unlike most of the research in this field, there is an extraordinary amount of data on what went on in the classrooms. Teacher usage of language in the classroom was tape recorded and then coded. Student usage of language was obtained by direct observation and coded on-site. As described, this part of the study design appears to be thorough, scientific, and quite impressive, although not without problems. Unfortunately, they use none of these classroom characteristics to explain achievement. All analyses are conducted of the nominal program type—immersion, early-exit, and late-exit—regardless of what actually went on in the classroom and how much it deviated from program type.

Only kindergarten through fourth grade data were collected for the immersion and early-exit programs, so even if they had wanted to they could not have compared the immersion/early-exit programs to the late-exit programs beyond 4th grade. Although the immersion program may have been too new to have more than four years of data, this is *not* true of the early-exit programs. Thus, their a priori decision to collect only K through 4 data for these programs was wrong, as anyone familiar with the research in this field can attest (see Baker & Rossell 1987; Rossell & Baker 1988). Students are not exited when they are supposed to from bilingual education programs. The authors of this study chose not to discuss this issue, and all descriptive data and analyses are presented as if students only

⁷ This is not the case for the early- or late-exit bilingual programs. The extent to which the distribution for immersion programs is more towards one extreme than the other can affect the findings in a sample this small.

enter a school district in kindergarten.⁸ We are never told how they deal with the new LEP students in their descriptive tables and statistical analyses.

Findings on program characteristics

Although there is a good deal of variation, the programs studied are, on average, different in the language of instruction both for the whole class and individual students. As shown in Table 1, teachers do appear to use very different amounts of English in the different programs and, not surprisingly, student usage of English follows teacher usage. These data indicate that the early-exit program is more like the immersion program in English language usage than it is like the late-exit program. The largest difference between programs is between the immersion and the late-exit programs with 64 percent less English usage in first grade by both teachers and students in the latter and approximately 40 percent less English usage as late as fourth grade by students.

Table 1
Average percentage usage of English by teachers and students

	TEACHERS				STUDENTS			
	Immer- sion	Early- Exit	Late- Exit	Late- Immer Diff	Immer- sion	Early- Exit	Late- Exit	Late- Immer Diff
K	98.5	65.8	9.3	-89.2	94.3	65.6	9.3	-85.6
1	97.3	69.1	32.9	-64.4	96.6	64.5	32.5	-64.1
2	98.2	74.5	30.3	-67.9	97.2	71.5	30.8	-66.4
3	99.0	80.3	50.6	-48.4	98.6	77.2	52.0	-46.6
4	99.8	97.3	55.3	-44.5	98.0	96.0	59.3	-38.7
5	*	*	63.6		*	*	65.3	
6	*	*	80.3		*	*	83.3	

*Data not collected for these grades. Source: Table 15, Ramírez, et al. 1991 (Vol I, p. 92), and Table 17, Ramírez, et al. 1991 (Vol I, p. 95)

These data show that, contrary to the researchers' criteria for each model, the early-exit programs, on average, use less than one

⁸ All of the target students, those directly studied, did enter in kindergarten. But the classrooms they are in consist of new LEP students. Fluent English proficient, and English-only students and the teachers behavior cannot help but reflect this.

hour a day of Spanish and the late-exit programs use Spanish less than 50 percent of the time by grade three. Moreover, these means obscure considerable variation within programs.

It is possible that the teachers are using more English with the target students—those who entered the program in kindergarten—than is reflected in these data which inadvertently may also include interaction with the new LEP students at each grade level. This appears to be the case with teachers' explanations in Spanish. Teachers in the *immersion* program increase their explanations in Spanish from 4 percent in kindergarten to 13 percent in first grade to 23 percent in second grade, and 17 percent in third grade (Ramírez, et al. Vol. I, p. 104). The only possible explanation for this is that the immersion teachers are having to deal with new LEP students each year and these tables include their interaction with them.

Although Ramírez, et al. decry what they see as a “passive learning environment” with too much seatwork and not enough oral practice, they did find that the “students exhibit a high level of task engagement” (above 87%). Given how heterogeneous these classrooms are—20-30 percent of these students are classified fluent-English-proficient (FEP) or English-only (EO) in both the early and late-exit programs in every grade beginning with kindergarten⁹—the researchers might not have observed this high level of task engagement if the teachers had spent more time in classroom discussion with their differently skilled students in a whole-class situation as the authors assert they should have done.

The theory behind bilingual education is that students will be able to participate in instruction at a more complex level, particularly in the initial stages, if their teacher uses their native tongue. There is very little difference, however, between the three types of programs in the percentage of students asking questions,¹⁰ commenting or initiating, or the percentage of instruction devoted to repetition, discussion, and drill activities (Vol. I, pp. 150-151). Thus, the

⁹ The authors do not tell us the basis for the categorization of students as LEP, FEP, or English-only. It is puzzling that the percentage of LEP students increases from one grade to the next in two grades of the immersion strategy and one grade of the late-exit program (Ramírez, et al. 1991, vol. I, p. 176). Does this mean this table includes new LEP students or that the scores of these students bounce around?

¹⁰ Through second grade there are more questions asked in the late-exit programs, but it is not clear what this means. Does it mean the students are more or less confused?

ability to use the student's native tongue does not cause the teachers to provide more complex instruction as measured in this study, nor does it keep the students more engaged in the tasks they were given (Vol. I, p. 156).

Another puzzling issue in this study is how the late-exit teachers can use so little English (9 percent in kindergarten and 33 percent in first grade) when between 10 to 15 percent of the students know only English in both the early-exit and late-exit programs? In addition, how can the early-exit teachers be using as much Spanish on a daily basis as observed when, on average, they are below the minimum proficiency level of 3.5 thought to be necessary to teach a language effectively (Vol. I, p. 184)? Is it possible that the teachers in the early- and late-exit programs used more Spanish when they were being observed than they used normally?

This suspicion is further aroused by the self-assessment ratings of the teachers (Vol. I, p. 198). About 20 percent of the teachers in the early-exit program responded that they do not speak Spanish. And from two to 10 percent of the teachers in the so-called developmental Spanish late-exit program in grades 1 through 4 responded that they do not speak Spanish. In addition, about 6 to 12 percent of the *late-exit* teachers across grades K-4 believe that use of Spanish among students should be discouraged (Vol. I, pp. 283-284) This is an odd viewpoint for a developmental program whose goal is to make the students fluent in two languages. Yet these teachers and their classrooms are included in the analysis of the effects of "bilingual" education.

Interestingly, almost 2/3 of the early-exit programs and 1/3 of the late-exit programs provide formal instruction in English reading in the first grade (Vol. I, p. 210). This provision of formal English reading is contrary to a basic tenet of TBE that reading should be taught in the primary language first and only when that is mastered should English reading begin. What we do not know, however, is whether some of the students already knew how to read in Spanish at a level that was satisfactory to the teachers or whether teachers simply determined they were not truly LEP and therefore could begin reading in English.

There is a difference in the amount of instructional time teachers feel they provide for different subject areas. The cost of being instructed in Spanish is a large decrease in the amount of time spent on English language arts and a small decrease in the amount of time

spent on math. This is shown in Table 2 where a minus sign means less time spent on a subject in the late-exit program.

From kindergarten through 3rd grade, the late-exit programs spend more than 1/3 less time on English language arts and, as shown in Table 3, the other subjects are taught in Spanish, not English. Therefore, the late-exit students are receiving about 2/3 less English time on task than the students in the immersion programs. In addition, they spend about five percentage points less time on math as a subject, even though it is instructed primarily in Spanish.

Teachers differ across the three programs in terms of whether they feel Spanish should be used for instruction. By first grade, only 1/3 of the early-exit teachers feel that Spanish should be used regularly for formal instruction, but 82 percent of the late-exit teachers feel that way. Thus, the early-exit TBE programs in this sample teach primarily in English.

As shown in Table 4, in the first four years of program participation (K-3), 28 to 50 percent of the students had not been reclassified as fluent-English-proficient and few had been mainstreamed into a regular classroom across all three programs. The authors do not discuss the reason why the students in the immersion and early-exit programs were not mainstreamed, but I suspect it is, at least in part, a function of the need to maintain a program through elementary school for the new LEP students who arrive at every grade level. Since the program must exist through grade 6 for these new LEP students, it is possible that the teachers keep the LEP students who entered in kindergarten in the program because they are needed as a critical mass to fill a classroom which must exist for those who do not speak English.

These data basically substantiate what Rossell and Baker (1988) reported—that most students are not exited from TBE programs even when they meet the criteria for exit. What we did not know at the time was that this would be the case for any whole-class, second language acquisition program in the U.S., including an all-English program. In addition, the exit criteria are biased toward keeping students classified as LEP (Baker & Rossell 1987; Rossell & Baker 1988). If the cut-off score for reclassification is the 36th percentile on the CTBS, a minimum of 36 percent of the LEP students will never be reclassified as FEP. Thus, a certain proportion will be incorrectly labelled LEP because their *true* score happens to be below the cut-off for being reclassified as FEP.

Table 2
Average percentage of instructional time for each academic subject

Grade	Immersion %	Early-Exit %	Late-Exit %	Late Immer. Diff	
K	English	63.6	35.1	25.8	-37.8
	Social Studies	7.0	6.9	8.6	1.6
	Mathematics	23.2	22.1	18.4	-4.8
	Science	6.2	6.1	7.3	1.1
	Spanish	0.0	29.9	39.9	39.9
1	English	60.2	33.5	23.2	-37.0
	Social Studies	7.5	8.0	7.2	-.3
	Mathematics	23.5	26.7	20.6	-2.9
	Science	6.6	7.6	7.4	.8
	Spanish	2.2	24.2	41.7	39.5
2	English	60.7	34.4	25.7	-35.0
	Social Studies	8.2	8.5	8.3	0.1
	Mathematics	24.1	23.1	18.7	-5.4
	Science	6.9	8.7	8.5	1.6
	Spanish	0.1	25.4	28.8	28.7
3	English	58.7	40.6	36.2	-22.5
	Social Studies	8.6	8.6	7.4	-1.2
	Mathematics	24.5	24.6	19.8	-4.7
	Science	8.2	8.1	7.8	-.4
	Spanish	0.0	18.2	28.9	28.9
4	English	52.5	51.1	43.0	
	Social Studies	18.0	14.6	9.1	
	Mathematics	17.5	20.9	19.5	
	Science	12.0	12.5	10.4	
	Spanish	0.0	0.9	18.0	
5	English	*	*	42.9	
	Social Studies	*	*	9.4	
	Mathematics	*	*	20.2	
	Science	*	*	9.4	
	Spanish	*	*	18.2	
6	English	*	*	41.0	
	Social Studies	*	*	11.7	
	Mathematics	*	*	18.3	
	Science	*	*	10.9	
	Spanish	*	*	18.1	

*Date not collected for these grades. Source Ramírez. et al 1991, Vol. I (p. 405).

Table 3
Average percentage Spanish used in instruction* by subject

Grade	Math			Science			Social Studies		
	Imm	EE	LE	Imm	EE	LE	Imm	EE	LE
K	0	42	87	0	9	93	0	9	93
1	0	25	83	2	14	65	2	18	84
2	0	17	78	0	22	57	0	19	78
3	0	27	53	0	23	48	0	27	55
4	0	0	40	0	0	43	0	0	40

*Spanish or primarily. Source: Ramírez, et al. 1991, Vol. I (pp. 219 -228).

Table 4
Percentage of students reclassified
or mainstreamed by end of third grade

Program	Reclassified %	Mainstreamed %
Immersion	66.7	25.6
Earl-Exit	72.0	16.9
Late-Exit	50.8	*

*No students mainstreamed because goal of program is Spanish language development. By 6th grade, however, only 79 percent had been reclassified as fluent English proficient. Source: Tables 165 and 166. Ramírez, et al. 1991, vol. I (pp. 372-373).

Although Congress keeps reauthorizing bilingual education, Title VII of the Elementary and Secondary Education Act, because they think it is the best way to teach children English, the district administrators clearly have another agenda in addition to teaching English. A primary goal of the bilingual education programs in this study is to maintain and improve the Spanish language ability of these children. We are, however, given no data on whether they are successful in this regard.

There are conflicting data regarding differences in the social status of the children in the three programs. Site administrators estimate the immersion program students to be of lower social class than those in the other programs (Vol. I, p. 319). The parent survey (Vol. I, pp. 351-352), however, indicates that the late-exit students have the lowest social class. The student data sheets, filled out by the teachers from student records, on the other hand, show no program difference in the percentage of students on free or reduced

lunch, except in kindergarten (Vol. I, p. 379), although more students are from poverty backgrounds in the immersion program (Vol. I, p. 284).

There are also more first grade students classified as LEP in the immersion programs than there are in the late-exit programs. But there is little difference in the use of Spanish in the family. About 78 percent of the parents in all three programs report speaking Spanish to their children and about 70 percent report their children speak Spanish to them. Since almost all of the target students had been living in the U.S. their entire childhood, and more than 3/4 had been to preschool where some English was spoken, they had all been exposed to English by the time they entered kindergarten. A certain percentage of these children were, thus, probably English dominant despite being classified as LEP on the basis of an oral English language proficiency test and/or a standardized achievement test.¹¹ Ramírez, et al. do not discuss this problem. They do show that in kindergarten only 30 percent of the immersion, 51 percent of the early-exit, and 52 percent of the late-exit students were classified as having no more than a beginning knowledge of English (Ramírez, et al. 1991, Vol. I, p. 388). Thus at least half of these kindergarten students were proficient in English according to their teachers. In first grade, the percentage not knowing English was even smaller in the immersion and early-exit programs, but about the same in the late-exit programs.

Although Ramírez, et al. conclude that there are roughly the same number of children with special needs across programs, my reading of their data is that there is a substantially larger percentage of learning disabled children in the immersion programs than in the two other programs in grades 1 and 2 (Vol. I, p. 381), and a larger percentage, as early as kindergarten, with such impediments to learning as a lack of English proficiency, a lack of motivation to learn and use English, a lack of parent involvement, a poverty background, underdeveloped cognitive skills and other problems (Vol. I, p. 384). More than 68 percent of the immersion first graders, but only 43 percent of early-exit and 17 percent of late-exit

¹¹ As discussed in Rossell and Baker (1988), oral English language proficiency tests will categorize, on average, 75 percent of English *monolingual* kindergarten students as limited-English-proficient. Standardized achievement tests will classify 36 percent of the English *monolingual* population at the 36th percentile or below, a common criterion for classification as LEP.

first graders are classified as learning disabled. Thirty-five percent of immersion second graders, 40 percent of early-exit but only 18 percent of late-exit second graders are classified as learning disabled (Ramírez, et al. Vol. I, pp. 381-382). These differences reverse, however, in third grade. Unfortunately, none of these data are used in the statistical analyses.

Their summary of the Volume I findings at the conclusion of that volume, are, with a few exceptions, judicious and accurate. One of these exceptions is their conclusion that “each program allocates approximately the same amount of instructional time in reading, language arts (collapsing instruction in English and Spanish), math, and social studies.” I strongly object to collapsing English language arts and Spanish language arts together in order to conclude there are no differences. Spanish language arts is *not* the same thing as English language arts, and combining them makes it look like there is no cost associated with bilingual education. The school curriculum magically fits one more subject without increasing the length of the day or reducing the other subjects! Because bilingual education advocates routinely combine Spanish and English language arts instruction and achievement in their description of programs and outcomes, many parents are, I suspect, unaware of the reduction in English language arts. Moreover, although the reduction in instruction is small in the other subject areas, it does exist, as shown in Table 3.

Ramírez, et al. also draw a few other conclusions that seem not to be supported by the data they present. They conclude, for example, that: “two important indicators of instructional quality suggest an advantage of late-exit programs. First, late-exit teachers assigned and correct homework more often than either immersion strategy or early-exit teachers” (Ramírez, et al. 1991, Vol. I, p. 426). The problem with this conclusion is that it assumes that correcting and assigning more homework is a unique advantage of late-exit developmental bilingual programs. This is certainly not true. Teachers in any kind of program can assign and correct homework. Rather than this being an advantage of late-exit programs, it should be seen as a systematic bias that might confound the analyses of achievement outcomes. If one program assigns more homework than the other, and if increased homework increases academic time on task, then that program’s academic outcomes will be inflated by the increased homework assigned and we will incorrectly attribute this to the program treatment rather than

the true cause. Since extent of homework assigned and corrected is not used as a control variable in any statistical analysis, this is another source of bias in this study.

In their summary chapter, Ramírez, et al. further criticize the characteristics of the second language acquisition programs they studied. They argue that it is a defect of all of the programs that not many EO students are mixed with the LEP students. Interaction with native speakers, they categorically state, is necessary for second language acquisition. They do not explain, however, how this might occur in a classroom taught predominantly in Spanish without its having a negative effect on the EO students. If having native English speakers as a model is an important factor in second language acquisition as Ramírez, et al. assert, then LEP students will have to be placed in English language classrooms because doing the reverse—putting EO students in bilingual classrooms—may negatively affect the academic development of the EO students. Even an immersion strategy program with instruction in English might negatively affect EO students if the teachers use less complex English to instruct the LEP students who comprise the majority of the students. The failure to consider the costs as well as the benefits of bilingual instruction unfortunately permeates most of the literature in this field.

Achievement outcomes

Volume II of Ramírez, et al. (1991) contains the research findings on the relative effectiveness of immersion, early- and late-exit bilingual education. It begins with a summary of the findings from Volume I which does not do justice to the richness of the data in that volume. Anyone who really wants to understand the characteristics of the three programs and their students must read the 458 pages of Volume I itself.

The authors lump the characteristics of the early-exit and late-exit models together and state that teachers in both these programs have native or near-native language skills in L1 and L2 (Vol. I, p. 184) when in fact the early-exit teachers had almost the same average below-minimum-proficiency level of Spanish ability (2.7 across all grades) as the immersion teachers (2.3 across all grades) in contrast to the late-exit teachers who had an average 4.2 across all grades. Thus, in many respects the early-exit program is more similar to the immersion strategy than it is to the late-exit program, and the authors

gloss over this by lumping the two bilingual education programs together.

Their three hypotheses are:

- Children in an early-exit program will have achievement¹² comparable to those in the immersion program;
- Late-exit children will demonstrate skills comparable to those in immersion and early-exit programs by grades 5 and 6;
- Late-exit children will be as proficient in content areas as immersion and early-exit children by grades 5 and 6.

Unfortunately, they only tested the first hypothesis. As indicated above, data were only collected from kindergarten to fourth grade for immersion and early-exit programs and no comparisons were made at *any* grade between late-exit and the other two programs.

Hierarchical linear model regression equations

The most important analyses of achievement outcomes are the hierarchical linear model (FILM) equations because they are the only ones that control for pre-program student and family differences. Table 5 summarizes the program treatment results of the achievement equations from kindergarten to first grade. The researchers made what I consider to be an unfortunate leap in what is otherwise a very careful progression of equations. They show no equations with variables controlling for pretreatment scores and student and family characteristics, but without the dummy school controls.

Table 5
K-1 Analyses for two program schools
(student pretest control)

Immersion v. Early-exit	Math	Language	Reading
Pred. 1st grade	N.S.	N.S.	Favors Early-exit (-12.4)

Source: Tables 46, 58, and 70, Ramírez, et al., vol. II (1991).

¹² The exclusive focus on achievement as an outcome is unfortunate because many advocates argue that the effects of bilingual education on self-esteem and motivation are more far-reaching and long-term than the effects on achievement.

School dummy variables (1=child in a particular school, 0=not) mask important policy relevant school practices that ought to be controlled for separately so the reader will understand what it is that is done in schools that affects achievement. A school cannot be duplicated, but school practices can.

Although Ramírez, et al. control for school and district in their equations, they give us no information on the extent to which schools vary by socioeconomic status and extent of English used. I suspect that schools do vary considerably in the extent of English used in the early- and late-exit programs and the social class of the LEP students. If that is the case, then controlling for school may eliminate some or all of the effect of program.

The fact that the early-exit students did significantly better on reading in first grade than the immersion students could be a function of the program, or of the preprogram characteristics of the students, or both. With only a one year time period—absolutely no bilingual education proponent believes the so-called facilitative effect of learning to read in the primary language is this immediate—and considerable self-selection bias in this sample, as well as school dummy variables that may eliminate or reverse the effect of the program, we just do not know the reason for this difference. Moreover, the early-exit program is not that different from immersion. There is certainly no basis for the authors' conclusion that "This is consistent with the primary language hypothesis that teaching students to read in their first language facilitates their learning to read in a second language" (Ramírez, et al. 1991, Vol. II, p. 126) since almost 2/3 of these students learned to read in English, almost 3/4 of instruction was in English, and most of these students were probably already proficient in English.

The authors' conclusion is further called into question by the analysis of grade 1-3 achievement for programs, summarized in Table 6, which shows no advantage in first grade reading for the same early-exit students. It does show, however, a significant advantage for the early-exit students in the *shape* of their growth curve from first to third grade. Because there is no advantage for overall annual *growth* from 1-3, this equation probably just shows the effect of increasing the extent of English used in instruction from 70 percent in grade 1 to close to 100 percent by the end of grade 3. Since the amount of English used in the immersion program is constant, they do not have the same upward curve. Rather they have a more constant growth. But the fact that this early-exit "catch-

up” does not occur in the other subject areas is puzzling and makes one wonder how valid any of these equations are.

Table 6 shows the immersion students doing better in language at the end of first grade. Again, this may be a program effect or it may be a result of self-selection bias. We just do not know. Because the kindergarten test is in a different metric than the outcome variable, it is not the best pretest, and cannot be part of these estimates for growth and curve.

Table 6
1-3 Analyses for two program schools
(All covariates and school dummy variables)

Immersion v. Early-exit	Math	Language	Reading
Pred. 1st grade	N.S.	Favors Immersion (11.0)	N.S
Growth 1-3	N.S.	N.S.	N.S.
Curve 1-3	N.S.	N.S.	Favors Early-exit (-17.5)

Source: Tables 89, 90, and 91, Ramírez, et al., Vol. 11(1991).

The K-1 and 1-3 analyses for one-program schools are shown in Tables 7 and 8. There is no difference among the programs in the K-1 analyses. In the 1-3 analyses there is a difference in the shape of the growth curve that favors early-exit programs. Again, because there is no advantage for overall annual growth this probably means only that the early-exit students are playing “catch-up.” As the percentage of English instruction increases, their English language achievement grows at an increasing rate not found in immersion programs where English is constant.

The fact that these differences flip flop from significant to insignificant and from favoring immersion to favoring early-exit on language and reading suggest that the findings are highly vulnerable to the sample of schools used in any particular analysis. In the language of statistical analysis, these findings are not robust, probably because, as the authors admit, there is more variation between schools than there is between programs. The small number of students in each of these sites in relation to the large number of variables may also contribute to the lack of robustness of these equations. In the most important analyses in this study—the two-program grade 1-3 equations with pretest data—there are only 86

immersion students and 28 early-exit students. In the one-program analyses for grades 1-3, there are 105 immersion students and 169 early-exit students.

Table 7
K-1 Analyses for one program schools
(all covariates and pretest)

Immersion v. Early-exit	Math	Language	Reading
Pred. 1st grade	N.S.	N.S.	N.S.

Source: Tables 120, 144, and 168, Ramírez. et al., Vol. II (1991).

Table 8
1-3 Analyses for one program schools
(all covariates and school dummy variables)

Immersion v. Early-exit	Math	Language	Reading
Pred. 1st grade	N.S.	N.S.	N.S.
Growth 1-3	N.S.	N.S.	N.S.
Curve 1-3	N.S.	Favors Early-exit (-14.1)	N.S.

Source: Tables 174, 175, and 176, Ramírez. et al., Vol. II (1991).

I suspect that not only is there considerable variation within program types in the most important policy variables, but these children were proficient or almost proficient in English when they entered kindergarten. They were, as shown in Table 9, close to the national norm by first grade. I find it hard to believe that children who truly were LEP when they entered kindergarten could come this close to the national norm (the 50th percentile) in that short a time. Either the true score of these students is well above the 50th percentile, which is not likely given their low social class, or after spending their entire childhood in the U.S., they were English proficient or close to it when they entered kindergarten, despite their classification.

Indeed, the small annual gains from first to third grade are not what LEP children are expected to achieve. Keith Baker, the original Department of Education project monitor for this study, calculated the annual gains of these students and found that if a LEP child entered kindergarten with a percentile score of 0 (because they

truly did not know English) they would have a percentile score of 1 by 6th grade at the rate of improvement of the children in this study!¹³ The only children who typically show such small annual gains in achievement are children who already know English. Thus this study may in reality be one of the effect of different language programs on children who already know English rather than the one the researchers thought they were doing—the effectiveness of different language programs on children who are LEP.

Table 9 also compares the predicted achievement scores of immersion and early-exit program students with all control variables in the equation. A minus sign in the column labeled “immersion-early-exit difference” in Table 9 means that the program favors early-exit. This table shows little difference between immersion and early-exit on math, language, and reading either in first or third grade.

Table 9
Adjusted achievement scores by grade and program in two program schools compared to national English-speaking norm

Grade	Nat'l Norm	Immersion	Early-exit	Immer.- Early-exit difference *
Math				
1	259	249	255	-6.00
3	364	331	334	-3.00
Language				
1	283	249	238	11.00
3	410	357	355	2.00
Reading				
1	263	228	232	-4.00
3	380	330	338	-8.00

*Minus sign favors early-exit; positive sign favors immersion. Source: Ramírez, et al 1991, Vol. II (p. 64. C9-C10).

Table 10 shows the predicted achievement scores for students in the one program schools with all control variables in the equations. Again there is little difference among the programs.

¹³ Personal communication with author, August 6, 1991.

Table 10
Adjusted achievement scores by grade and program in one program schools compared to national English-speaking norm

Grade	Nat'l Norm	Immersion	Early-exit	Immer.- Early-exit difference *
Math				
1	259	258	256	2.00
3	364	348	352	-4.00
Language				
1	283	256	257	-1.00
3	410	393	384	9.00
Reading				
1	263	233	240	-7.00
3	380	353	354	-1.00

*Minus sign favors early-exit; positive sign favors immersion. Source: Ramírez et al. 1991. Vol. II (p. 64, C10-C11).

Table 11
Unadjusted weighted achievement scores by grade and program for immersion and late-exit programs compared to national English-speaking norm

Grade	Nat'l Norm	Immersion	Early-exit	Immer.- Early-exit difference *
Math				
1	259	250	250	0.00
3	364	336	347	-11.00
Language				
1	283	252	261	-9.00
3	410	378	360	18.00
Reading				
1	263	230	239	-9.00
3	380	344	336	8.00

*Minus sign favors early-exit; positive sign favors immersion. Source: Ramírez et al. 1991, Vol. II (p. 60-61, 64, C3-C7).

Table 11 shows the unadjusted weighted average achievement scores comparing the immersion and late-exit students. Adjusted scores could not be used to compare the two programs because the adjusted scores for each were derived from different equations and different comparisons (immersion versus early-exit and late-exit sites against each other). Thus, I have used here the unadjusted averages weighted by the number of students in each program. The unadjusted scores, however, do not control for pretreatment differences. Nor should the first grade score be considered an adequate pretest since it was obtained after two years of the program and we do not know whether it reflects program or student differences.

This table shows that the immersion students start off doing worse in reading and language at the end of the first year, but end up doing better than the late-exit students. The immersion students start off the same in math, but end up doing worse. What does this mean? It could mean that the immersion students are not as smart as the late-exit students (we are never given the kindergarten test results so we do not know), but that the effect of being taught in English eventually gives them an advantage over the late-exit students in English language and reading, but not in math because it is less language based. It might also mean that the effect of being taught partly in one's primary language gives one an advantage initially, but that over time the effect of less English takes its toll. It could also mean nothing since most of the coefficients that produced these predicted scores were not statistically significant.

Trajectory analysis of matched percentiles graphs

Although the authors never directly compare late-exit and immersion programs, they do devote an entire chapter to an analysis of graphs, called TAMP (Trajectory Analysis of Matched Percentiles), that chart the progress of student cohorts from one grade to the next in each type of program. They very judiciously caution the reader that, although the relative gains of two or more groups can be evaluated with these graphs, "the graphs cannot identify the source of the differences; for example whether differences stem from student ability or from the educational program" (Vol. II, p. 56). They then proceed to violate their own admonition! They conclude at the end of this chapter:

When late-exit program sites are examined more closely, it appears that those sites that provided their students with the most primary language instruction consistently grew faster than this norming population (Ramírez, et al. 1991, Vol. II, p. 641).¹⁴

In the final chapter of Volume II, which will be the only thing that many people read, they conclude:

The TAMP curves suggest that students in site E, who were provided with substantial instruction in their primary language and a slow phasing in of English instruction over time, consistently realized the greatest growth in mathematics skills (Ramírez, et al. 1991, Vol. II, pp. 653-654).

What they neglect to tell the reader is that the students in site D, which had *less* Spanish instruction than those in site E, did better in math, language, and reading by their criterion—extent of improvement in TAMP scores (which is not a valid criterion because there are no controls for pre-program student differences) than those in site E, which had the most Spanish instruction.

The most glaring omission in their discussion of these graphs and in the rest of the report is a failure to inform the reader that standardized achievement tests are designed for English monolingual students so that students who make grade level progress will leave with the same percentile score that they entered with. This is *not* the case with LEP students. They are supposed to make *large* gains relative to the English monolingual population norm because they are starting from a point where they have low scores because they know little English. If the tests have any validity, their later scores should be significantly higher than their earlier score when they knew little English. The fact that, on average, the students in these studies, except in one district, did not do this suggests: 1) there is something wrong with *all* the programs, or 2) the students were fluent or near fluent in English in first grade and this is why their curves were similar to the English-monolingual population.

¹⁴ The authors conclusions regarding the utility of using more Spanish in the late-exit programs rests on 16 students with a pretest in site D, 82 in site E, and none in site G! The TAMP curves, which have no controls for other variables, include 35 students in sites with data from grade 1 through 6, 21 in site D, and 98 in site E.

In addition, my interpretation of the slope of the 209 TAMP curves for the different programs is quite different from Ramírez, et al.'s interpretation. Just looking at these curves, I would say that the relationship between Spanish instruction and achievement is curvilinear with the program having the least Spanish doing the best and the program having the most Spanish doing better than the one with an intermediate amount of Spanish (early-exit). Of 73 late-exit graphs, 11 percent show a decline relative to English monolingual students. Of 69 early-exit graphs, 20 percent show a decline relative to English monolingual students. Of 67 immersion graphs, however, only 1 percent show a decline relative to English monolingual students. While this does not include controls to eliminate the differences between students and schools, it certainly demonstrates the lack of evidence for Ramírez, et al.'s conclusion that primary language instruction appears to produce greater gains for LEP children. It "appears" the opposite to me.

Conclusion

This is an impressive study in terms of the massive quantity of data collected from the administrators, teachers, parents, and students over four years, and the classroom observations. The authors are clearly a technically skilled group of researchers. Unfortunately, although Volume I presents a wealth of informative, carefully collected and constructed data, little of it was used in what most people would consider the *raison d'être* of this study—the analysis of program outcomes. As a result there is much that is left to be done.

One conclusion that could be drawn from this study as it currently exists is that in terms of academic achievement (tested in English), it does not matter whether language-minority children who may be proficient in English are instructed in Spanish or English. Thus, the analysis provides no support for the "facilitative"¹⁵ effect

¹⁵ Ramírez, et al. (1991) may be using this word differently than it is used in the bilingual education literature. As they seem to use it the facilitative effect of primary language instruction enables a student to cancel out the negative effect of not being instructed in English so that he or she eventually has the same achievement as one would have if instructed in English. Among bilingual education theorists, however, the facilitative effect of primary language instruction produces greater achievement gains in English than if one had been instructed in English from the beginning.

of primary language instruction,” but neither is there support for the principle that academic time on task is the greatest predictor of achievement in any subject. Large deficits in English language instruction over several grades apparently make little or no difference in a student’s achievement.

It could also be inferred from this study that we ought to get rid of *all* special language acquisition programs, since they are expensive and they appear to be no better for LEP students than regular classroom enrollment with ESL pull-out. I infer this from the fact that the most common finding of the studies comparing ESL pull-out to transitional bilingual education is that there is no difference between the two (Rossell & Ross 1986; Rossell 1990). If there is no difference between immersion and bilingual education and no difference between bilingual education and ESL pull-out, it is not unreasonable to assume that there is no difference between immersion and ESL pull-out, although they were not compared directly in this study. Given the problems with this study and the lack of robustness of the findings, however, I would not place much confidence in these conclusions.

The following analysis of the existing data¹⁶ needs to be performed before I would believe these findings on the effect of language of instruction on the achievement of LEP children. The analysis that should be done would use the percentage of English used in instruction in a subject (math, language, or reading) (rather than the “nominal” program type, immersion, early-exit, and late-exit) by classroom as the independent variable predicting achievement in the *same* subject area controlling for student pretest data, student school behavior, family characteristics, and district, school, and teacher characteristics unrelated to the program which, if not controlled, might be confused with a program effect. These teacher variables should include the extent of homework assigned and corrected,¹⁷ the extent of language acquisition training, the

¹⁶ There are additional data - such as 5th and 6th grade achievement for the immersion and early-exit students, and self-esteem data that should have been collected, but it is too late for that.

¹⁷ Parent assistance with homework, however, is probably a real program effect favoring the late-exit programs. It makes sense that parents who are themselves limited in English and proficient in Spanish will be able to help more with homework when it is in a language they can understand. Therefore, this variable

extent of Spanish language ability, and other characteristics thought to make a good teacher which are not unique program effects. Students in all programs (immersion, early-exit, and late-exit) should be directly compared to each other in this analysis,¹⁸ although the nominal program type would not be a variable nor would there be dummy school variables that mask important policy relevant variables.

Finally, this analysis should be conducted separately for students who were truly LEP in kindergarten and those who are actually FEP, but classified as LEP. The extent of misclassification in any particular school can be quite large. Knowing this, researchers should take special care to determine those who are truly LEP. One way to do this with the information collected in this study is to assume that any kindergarten student who is able to take the Tests of Basic Experiences (TOBE) in English is potentially FEP regardless of their later score on the CTBS. The teacher's rating of each individual student's proficiency in English should be a factor since research shows that teachers are the equal of foreign language experts and are better judges of a student's English language proficiency than are standardized tests (see Baker & Rossell 1987; Rossell & Baker 1988). In addition, one could compare a student's Spanish language proficiency¹⁹ and their English language proficiency. Those who score lower in Spanish than in English are potentially FEP regardless of their English language score. In addition to these criteria, students classified as LEP could be divided into those who score within a few points of the cut-off score for classification and those who score significantly lower. None of these criteria by themselves are fool-proof, but taken collectively might be able to distinguish children who are truly LEP when they

should *not* be controlled because doing so would eliminate a unique advantage of the late-exit program that might potentially affect academic outcomes.

¹⁸ I do not understand why the same technique used to compare the early-exit and immersion programs in different schools and the late-exit programs in different districts could not be used to compare the late-exit to the immersion programs in different districts since, as the authors found out, there is more variation between schools than between districts.

¹⁹ None of the Spanish language achievement data that were collected are described or analyzed in this report.

enter these programs from those who are probably FEP, but for various reasons are classified as LEP.

The same analysis should also be conducted on other outcome variables on which there are data, such as the number of absences. Although Ramírez, et al. treat this variable as an independent student characteristic, it can also be an outcome of a particular educational climate. Since there are those who argue that bilingual education promotes self-esteem and reduces drop-out, but these data were not collected, the number of absences might be a reasonable alternative.

In conclusion, this study is, in many respects, an impressive piece of research and analysis. There is, however, much that needs to be done before I am willing to accept the findings as real. As Ramírez, et al. conclude:

There was a diversity of patterns in the amount of English used among early-exit sites in grades kindergarten through second grade.. .The one late-exit site that more closely resembled an early-exit program is more of a concern.. .The effect of these differences in patterns of English among the early-exit sites as well as among the late-exit sites would best be addressed in future analyses of operational programs (Ramírez, et al. 1991, Vol. II, p. 35).

But this is only the beginning of any future reanalysis.

References

Baker, K. & C. Rossell 1987. An implementation problem: Specifying the target group for bilingual education. *Educational Policy*. 1 (2), 249-270.

Danoff, M., G. Coles, D. McLaughlin & D. Reynolds. 1978. *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program* [Year Two Impact Data]. Palo Alto, CA: American Institute for Research.

Danoff, M., B. Arias, G. Coles & B. Everett. 1977. *Evaluation of the impact of ESEA Title VII Spanish/English Bilingual Education Program*.

Ramírez, J., Yuen, S., 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., D. Ramey & 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.

Rossell, C. 1990. The Effectiveness of Educational Alternatives for Limited English Proficient Children. In G. Imhoff (Ed.), *The Social and Cultural Context of Instruction in Two Languages: From Conflict and Controversy to Cooperative Reorganization of Schools.* New Brunswick, NJ: Transaction Publishers.

Rossell, C. & K. Baker. 1988. Selecting and Exiting Students in Bilingual Education Programs. *Journal of Law and Education.* 17(4), Fall, 589-624.

Rossell, C. & J. Ross. 1986. The social science evidence on bilingual education. *Journal of Law and Education.* 15, 385.