

## Appendix: Comments on a Recent Critique

---

Basing their conclusions on an analysis of 72 studies that they considered to be methodologically acceptable, Rossell and Baker (1996) conclude that the research evidence does not support transitional bilingual education as a superior form of instruction for limited English proficient children (p. 7). I was not able to re-examine all the studies Rossell and Baker cited, because most were unpublished reports. I did, however, read all of the studies they cited that appeared in the professional literature, and found numerous problems with Rossell and Baker's conclusions. I examine here studies in which submersion and "immersion" are claimed to be superior to bilingual education. I then discuss those studies Rossell and Baker categorized as "unacceptable."

### **Submersion versus Bilingual Education**

Rossell and Baker conclude that in the case of English reading comprehension, transitional bilingual education was superior to submersion in 22% of the studies, worse in 33%, and there was no difference in 45% of the studies (60 studies were examined). Of the 20 studies in which submersion was claimed to be superior to bilingual education, two were in the published professional literature and I was able to get information on one other.

Moore and Parr, 1978: This study examined children in four programs: maintenance, transitional, minimal and English-only, and concluded that the latter group scored significantly higher than the others on tests of English reading. We are, however, given no details whatsoever on what went on in the bilingual classes; all we have are labels. In addition, the duration of the study was short: Moore and Parr's oldest subjects had just finished grade two. It typically takes longer for bilingual programs to show a positive impact on English language tests. In addition, no raw data is provided, so it is difficult to tell what the real effect of each program was.

Curiel, Stennina and Cooper-Stenning, 1980: In this study, seventh graders who had been in bilingual education were compared to comparison students who had not. While Rossell and Baker classify this study as showing submersion to be superior, this is not what Curiel et. al. report. For reading comprehension

tests given in grade six, non-bilingual education students were significantly better. But at grade seven, there was no significant difference between the groups, although the comparison students were slightly better (mean = 6.35, compared to 5.98). In addition, students in bilingual education had significantly higher grade point averages and fewer of them were retained (30 out of 90 controls had been retained one year in elementary school, compared to 11 out of 86 bilingual education students). Finally, Curiel et. al. note that the bilingual program was used as a remedial program for some students who were previously placed in the monolingual English program.

Rossell and Baker complain that no one looks at the future educational success of graduates of bilingual or immersion programs (p. 41). Curiel and his associates attempted to do this. In Curiel, Rosenthal, and Richek (1986), students who were studied in Curiel et. al. up to grade seven, described just above, were followed to grade 11. (Rossell and Baker cite this study, but classify it as methodologically unacceptable.) In the follow-up, Curiel and associates report that students who had been in bilingual programs outperformed comparisons on all measures. While 12 of 90 comparison students had been retained in grades seven, eight and nine, only 4 out of 82 bilingual education students had been retained. Fewer bilingual education students dropped out of school (23.5%, compared to 43%). Bilingual students had higher grade point averages, but the difference was not significant (note that those who dropped out probably had lower grades.)

El Paso: Rossell and Baker include El Paso evaluations as studies showing submersion to be more effective than bilingual education. This is a bizarre analysis: neither program was a submersion program. Two programs were compared in El Paso. One, labelled "bilingual immersion" by the El Paso Unified School District, was clearly bilingual education. It contained a "native language cognitive development component" (NLCD), described by the El Paso Independent School District (1989b) as follows:

"NLCD is taught for 60 to 90 minutes per day. The objective of this component is to develop concepts, literacy, cognition, and critical thinking skills in Spanish. It is during this period that instruction and student-teacher interaction are entirely in Spanish. The more demanding content area concepts are also introduced during NLCD, particularly in the first grade (p. 54)."

This program employed the Natural Approach for ESL, a whole language approach to language arts, and sheltered subject matter teaching.

Thus, "bilingual immersion" in El Paso combined instruction in the first language with comprehensible input-based methodology, similar to the gradual exit variable threshold plan described in chapter 2.

The second program was also considered to be bilingual education, but differed in some important ways. Referred to as SB 477, it used a skills oriented approach:

It must be understood that BIP (bilingual immersion program) is not an English version of the SB 477 instructional program. SB 477 is built on a philosophy that advocates traditional concepts of teaching language ... SB 477 focuses the child's attention on the details of language such as phonetic sounds and grammar rules" (p. 9).

While bilingual immersion used whole language and Natural Approach activities, the most commonly used materials in SB 477 were basal texts and workbooks (El Paso Independent School District, 1987, p. 18). According to a 1989 report, whole language and comprehensible input-based methodology had been gradually introduced into SB 477 from 1985 to 1987, but observations indicated that the changes had not been fully implemented by SB 477 teachers (El Paso Independent School District, 1989a, p. 10).

Gersten, Woodward, and Schneider (1992) published a detailed report on the El Paso programs, limiting their analysis to students who had been in either program continuously. Gersten et. al. confirmed that the bilingual immersion students excelled in grade four in all aspects of academic performance, but by the seventh grade, no significant differences were found (p. 13). Far more bilingual immersion students, however, had been placed in the mainstream at grade 6, and more bilingual immersion teachers were confident about their students' eventual success in the mainstream.

Neither program was particularly successful, however. Sixth graders did acceptably well on the Language and Math Subtests of the Iowa Test of Basic Skills, but did poorly on the Reading Comprehension test, with bilingual immersion students scoring around the 23rd percentile and SB 477 students scoring at the 21st percentile. Vocabulary scores were even lower. The students themselves said that their hardest subjects were social studies and language arts, those which demand the most competence in academic English (pp. 25-26). In addition, even though bilingual immersion teachers were more confident of their students' eventual success in the mainstream, only 73% thought they would succeed in the regular program, and only 45% of the SB 477 teachers predicted success for their students. Gersten et. al. note that "These levels of performance, sadly, are typical for low-income Hispanic students in the junior high school years" (p. 29). This may be true, but as argued in chapter 8, I think we can do much better.

All we can conclude from the El Paso program is that a well-designed bilingual program was better than a less well-designed program up to grade four, with

no differences (or a slight edge for bilingual immersion) by grade seven. Such studies will help us decide among bilingual education options, but do not address the issue of bilingual education versus submersion.

**Berkeley:** Rossell and Baker include a study by Rossell (1990) of LEP performance by children in Berkeley as showing that submersion plus ESL was, in one case, superior to bilingual education, and in another that there was no difference in English reading. The programs Rossell compared were labelled bilingual education and pull-out ESL, but no description of Berkeley's bilingual education program was provided, other than the fact that it was labeled bilingual education and that instruction was in Spanish 30 to 50 percent of the time.<sup>1</sup>

### **Immersion versus Bilingual Education**

The term immersion has been used in a number of different ways. Here, I will focus on just two uses.

**Canadian-style immersion (CSI):** As is well-known, CSI is a program in which middle-class children receive much of their subject-matter instruction through a second language. Efforts are made to make sure the language they hear is comprehensible. Children in these programs learn subject matter successfully and acquire a great deal of the second language.

Consideration of the principles of bilingual education presented in chapter 1 leads to the conclusion that CSI is similar, if not identical, to bilingual education. Children in CSI receive comprehensible input in the second language and develop literacy and subject matter knowledge in their first language, both outside of school and in school; children in CSI are typically middle class, and do a considerable amount of reading in English outside of school (suggested by Cummins, 1977, and confirmed by Eagon and Cashion, 1988). Even in early total immersion programs, a great deal of the curriculum is in English, with English language arts introduced around grade two. By grade six, half the curriculum of early total immersion is in English. Most important, the goal of CSI is bilingualism, not the replacement of one language with another.

**Structured immersion (SI):** As described by Gersten and Woodward (1985), SI has these characteristics:

- (1) Comprehensible subject matter instruction.
- (2) Use of the first language when necessary for explanation, but this is kept to a minimum.

- (3) Direct instruction of grammar.
- (4) Pre-teaching of vocabulary.

While the first two characteristics are supported in the research literature, there is little evidence supporting the efficacy of direct grammar instruction (Krashen, 1994) and pre-teaching vocabulary has not been found to be consistently effective (Mezynski, 1983).

Rossell and Baker (1996) claim that immersion was more effective than bilingual education for English reading in eight studies. Information concerning six of these studies was available.

Barik, Swain and Nwanunoki (1977) and Barik and Swain (1978) are studies of Canadian-style immersion with French as the target language in which early total immersion is compared to partial immersion. In partial immersion, there is less teaching in French; from the beginning, some subjects are taught in English and some in French.

While Rossell and Baker are not fully explicit concerning why these studies were included, the idea seems to be that early total immersion is similar to all-English immersion for LEP children and partial immersion is similar to bilingual education. Since Barik et. al. and Barik and Swain show that children in early total immersion acquire more French than children in partial immersion programs, immersion, it is concluded, is better than bilingual education.

But Canadian early total immersion is not the same as an all-English immersion program for LEP children. In fact, both versions of Canadian-style immersion under consideration here, early total and partial immersion, are quite similar to bilingual education. As noted earlier, much of the CSI curriculum is in the first language, English, and children in these programs typically come to school with a great deal of literacy development in the primary language. Since children in both programs come to school so well-prepared, it is reasonable to expect that more exposure to the second language, French, will result in more acquisition of French, because what is heard and read is mostly comprehensible.

Many LEP children in the United States, however, do not come to school with these advantages. An all-second language curriculum will be much less comprehensible to them, even if carefully sheltered. While sheltering will clearly help, supplying background knowledge and literacy in the first language is a sure way to ensure that instruction in English will be comprehensible.

Rossell and Baker are clearly aware of this argument. They point out, in defense of their position, that Canadian-style immersion programs have worked for working class students as well as middle class students. A few reports of Canadian-style immersion programs for working class children have been published (e.g. Holobrow, Genesee, Lambert, Gastright and Met, 1987). While these children have done well, evaluations so far have been limited to grade two and below. Also, as Genesee (1983) notes, none of these children can be said to come from destitute or 'hard-core' inner-city areas (p. 30).

We thus know very little about how well working class children do in second language immersion programs and nothing about how well under-class children would do. What we do know is that children of lower socioeconomic background experience less print outside of school and that the richness of the print environment is related to literacy development (evidence cited in chapter 3.)

Bruck, Lambert and Tucker (1977) compared children in total immersion (CSI) to native speakers of French, and thus has no bearing at all on the issue of bilingual education versus immersion.

Genesee, Holobrow, Lambert and Chartrand (1989) is a comparison of fifth graders in partial immersion, early total immersion, and the performance of English-speaking children in a French school designed for native speakers of French. This third group, however, had methodology that was quite similar to the students in early total immersion, and the population of the school was largely English-speaking. In fact, Genesee et. al. refer to it as super-immersion. Super-immersion students and early total immersion students performed similarly on nearly all measures.

An additional group of a small number of English speaking students in a French school with fewer English-speaking classmates was compared to the super-immersion students; performance was similar, with the submersion students doing better only on oral tests. Again, it is not clear why this study is included and how it is supposed to show that structured immersion is better than bilingual education. In my analysis, all groups had the benefits of bilingual education, with first language development coming from school and from home. We also cannot use the French school-immersion comparison to claim that submersion is equivalent or superior to bilingual education because we have no idea how the English speakers were dealt with in classes in the French school, that is, how much comprehensible input in French they received, in school and outside of school (e.g. tutors).

In Gersten (1985), it is claimed that more students in structured immersion scored at or above grade level on standardized tests than children in bilingual education. Gersten's study meets the methodological criteria set out by Rossell and Baker (see below), but is full of other problems.

- Sample size: Gersten compared only 28 structured immersion students to 16 students in bilingual education. Because Gersten does not provide actual scores for the bilingual group, we have no idea how close to grade level these students were (one cohort of SI students scored at the 64th percentile, the other at the 65th percentile). It is possible that if only a few students had slightly different scores, the results could have looked very different.

- Duration: Gersten examined student performance at the end of second grade. Gersten included follow-up data on the SI students, and they did very well, with one cohort scoring at the 65th percentile in reading at the end of grade four and the other scoring at the 78th percentile in reading at the end of grade three. Each cohort, however, consisted of only nine children!

- Control for socio-economic factors: As noted earlier, SES has a powerful effect on school performance. Gersten informs us that his SI students were low income and that the school they attended had a high proportion of low income, low achieving students, who became eligible for Title I funds (p. 188) but provides no supporting evidence of any kind. Nor is SES information provided about the comparison students in the bilingual education program. Comparison children did not attend the same school the SI children attended. (We do not even know the name or location of the school or district studied. All we know is that the school is on the West Coast.)

- Lack of information about the bilingual education program: We are only told that comparison students participated in bilingual education programs in the district. We have no idea what the quality of the program was, what methodology was used, etc.. In addition, the comparison (bilingual education) students in cohort II included two speakers of Korean, two speakers of Vietnamese, and two speakers of Samoan or Thai. This implies that this mysterious district provides full bilingual education programs in all of these languages at least up to grade two. I know of no districts that are able to do this. Gersten did not include information on the linguistic background of the bilingual education students in the other cohort. Such data, he states, was unavailable.

Finally, Gersten notes that the number of LEP students in the school he studied is small; his analysis included all LEP children who were in the program for at least eight full months (p. 190). These children, thus, were among many English-speaking peers. We do not know what the linguistic environment was like for the children in the bilingual education program.

Everything is wrong with this study.

Pena-Hughes and Solis, (1980) is unpublished, but it is discussed in several published papers. It is a comparison of two programs in McAllen, Texas. While

Rossell and Baker label these programs "immersion" and "bilingual education." Willig (1985, 1987) classified the immersion group as bilingual education, noting that the immersion group had instruction in English in the morning and instruction in Spanish reading in the afternoon. In addition, the explicit goal of the immersion program was bilingualism—development of both languages.

Also, the group Rossell and Baker label "bilingual education" did not, apparently, have a good program. According to an article in the *Wall Street Journal* (Schorr, 1983), classes were conducted partly in Spanish and partly in English, suggesting concurrent translation, a method shown to be ineffective (Legarreta, 1979). What apparently happened in McAllen is that children in a good bilingual program outperformed children in a poor bilingual program.

### **"Unacceptable" Evidence for Bilingual Education**

In this section, I review several published studies that were either not cited by Rossell and Baker or classified as "unacceptable". My conclusion is easy to state: The "unacceptable" or omitted studies either support bilingual education or are irrelevant. The criteria Rossell and Baker used to exclude studies are the following (pp. 14-15):

1. "The study did not compare program alternatives or assess educational outcomes." My interpretation of this criteria is that the study had to compare bilingual education to something else (the title of their paper is "The educational effectiveness of bilingual education." In their tables, they include, however, a comparison of "immersion" versus "ESL", which does not involve bilingual education and also include a comparison of two versions of bilingual education: transitional versus maintenance<sup>2</sup>).
2. "The study did not use randomly assigned students and made no effort to control for possible initial differences between students in different programs." This criteria, we will see, is one that is frequently violated in the "unacceptable" studies. I will argue, however, that the results of these studies are too strong to be ignored and that there is reason to hypothesize that this violation is not as serious as Rossell and Baker suggest it is.
3. "The study did not apply appropriate statistical tests."
4. "The study used a norm-referenced design." These studies typically compare bilingual education students to national norms. Rossell and Baker argue that this is not valid because one would expect limited English proficient children to show dramatic gains once they acquire some English and can show their



true competence on tests. Nevertheless, strong progress in comparison to norms is certainly consistent with the hypothesis that bilingual education is effective.

5. "The study examined gains over the school year without a control group." Without question, gains seen without a control group are much less convincing. But they are suggestive, especially when they are far beyond expectations.

6. "The study used grade-equivalent scores." While imperfect, outstanding results with grade-equivalent scores should not, in my opinion, be ignored.

7. "The study compared test results in different languages for students in different programs." All test results to be discussed in this paper are tests given in English.

8. "The study did not control for the confounding effect of other important educational treatments that were administered to at least one of the groups, but not all of them." If the bilingual group did better, but had other treatments, one cannot say for sure that bilingual education was responsible for the advantage.

### **"Unacceptable" but Suggestive Studies**

In this section, I survey studies that were classified as unacceptable by Rossell and Baker but that, nevertheless, provide interesting information. As we will see, the requirement most frequently failed of the eight listed above is number 2, failure to randomize or control for possible initial differences. The studies are presented chronologically:

Kaufman (1968) is classified both as "acceptable" and "unacceptable" in Rossell and Baker, and is listed as showing TBE to be superior to submersion. Kaufman compared Spanish-speaking junior high school students who were randomly assigned to one of two groups in grade seven. The experimental group had instruction in Spanish reading. This group consisted of two subgroups, one having Spanish reading instruction for four times per week for two years, and the other for only one year. The control groups had art, music and health education during the time the experimental group had Spanish reading. Each experimental subgroup had its own control group from the same school. Both groups "received equivalent instruction in English" (p. 523).

Table A.1 presents test results for English vocabulary and reading (Durrell-Sullivan Reading Capacity and Achievement Tests) adjusted for pre-test differences. In every case, the experimental groups did better, but in no case

was the difference statistically significant. Effect sizes were low in some cases, and modest in others. Thus, this study could be classified as being pro-primary language instruction or "no difference," depending on the analyst's view of statistical significance.

Kaufman's study failed one of Rossell and Baker's requirements for acceptability; test scores were presented as grade equivalent units, not as raw scores.

Table A.1.  
Reading and Vocabulary Scores

	<i>n</i>	<i>means: vocabulary</i>	<i>reading</i>
Spanish reading: two years	31	5.51	5.65
Control	19	5.29	5.53
effect size		.31	.16
Spanish reading: one year	20	4.86	4.87
Control	25	4.83	4.57
effect size		.04	.47

*effect size computed by SK*

Rosier and Farella (1976) and Vorih and Rosier (1978) reported that children in a Navajo-English bilingual program at the Rock Point school had better attainment in English when compared to English-only schools on the Navajo reservation, and did better than previous cohorts at Rock Point who did not have bilingual education. The study did not utilize random assignment and did not attempt to control for pre-test differences. Rosier and Farella (1976), in fact, note that Rock Point averages had been higher than those in the other schools since 1963-64. They were still, however, two years below national norms. In addition, the Rock Point studies utilized grade-equivalent scores, another violation of the Rossell and Baker requirements.

The Rock Point scores are nevertheless very impressive: Fifth graders in Rock Point who had had bilingual education scored 5.0 in 1975 and 5.4 in 1976 (compared to previous cohorts' 3.9 and 3.8) and sixth graders in 1976 scored 6.6 in reading comprehension (compared to a previous cohort's 4.7). Clearly, something good was happening at Rock Point.

Baker and de Kanter (1983) have other criticisms of the Rock Point data. They note that both the bilingual education students at Rock Point and previous cohorts experienced a large jump in test scores between grades 2 and 6; this increase, thus, can not be attributed to bilingual education. Even so, the Rock Point attainments in grades five and six are high. In addition, Baker and de Kanter point out that different versions of the SAT test were used in different years. Again, this is a flaw, but again, the scores of the fifth and sixth graders are a clear improvement. Finally, Baker and de Kanter (1983) point out that some students without a full experience in bilingual education may have been included in an analysis of combined scores. The high attainments of the fifth and sixth graders cited here are not a result of this analysis.

Ferris and Politzer (1981) compared English language competence and school success in two groups of Spanish-speaking junior high school students. One group was born in Mexico and had completed at least three years of education in Mexico, in Spanish. The second group was born in the USA and had had all schooling in English. Ferris and Politzer reported no difference between the groups on an essay written in English, "except for minor differences" in favor of the US-born group on points of grammar (verb inflections, pronoun agreement; there was no difference between the groups for paragraph development, sentence boundaries, article agreement, possessives, clauses per T-unit, and average T-unit length).

Of great interest is the finding that the group that had had some education in Mexico had significantly higher grades in English (mean 3.17 out of 4 compared to 2.53), reported that they tried harder to get good grades, and reported more discussion with teachers about school work.

Gale, McClay, Christie, and Harris (1981) Gale et. al. compared Australian aboriginal children in all-English schools and children who had bilingual education (Gapapuyngu). The bilingual model presented by Gale et. al. did not utilize translation, and gradually shifted instruction into English, beginning with math and English literacy. Gapapuyngu language arts was maintained until grade four. When tested at grade five, there were no differences between the groups in English vocabulary and story retelling (fluency), and the English-only children were better on a cloze test. By grade seven, however, the bilingual education group was far better on tests of fluency, on a cloze test, on English composition, on tests of subtraction, multiplication and division, and tended to be better in reading (table A.2).

Table A.2.  
Grade Seven Results: Gapapuyngu Study

Test	English only	Bilingual	Effect size
Vocabulary	51.5	49.5	.42
Fluency	111.1	132.7	.53
Reading	6.70	7.18	.40
Cloze	24.0	52.5	1.00
Essay	8.8	12.9	1.52

Effect sizes calculated from *t*-values in Gale et. al.  
From: Gale et. al. (1981).

As Gale et. al. note, there were flaws. Rossell and Baker's criteria 8 was violated: English-only controls were previous cohorts, and Gale et. al. point out that other curricular developments had been put in place and that the community was "becoming more Europeanized" (p. 301), with greater exposure to English. In addition, estimates of validity were not done for the locally developed tests. Nevertheless, the results are very strong.

Lofgren and Ouvinen-Bierstam (1982) compared the achievement of Finnish-speaking students living in Sweden who participated in a bilingual program to other immigrant children and native speakers of Swedish. Table A.3 presents results at grade 3:

Table A.3.  
Bilingual Education in Sweden

Test results at grade 3	Finnish children	Other immigrant	Swedish children
n	32-34	29-46	33-62
Swedish standardized achievement test	1.9	1.9	2.3
Swedish	2.7	2.4	2.9
Mathematics	2.9	2.7	3.0

from: Lofgren and Ouvinen-Bigerstam, 1982

This study was classified as "unacceptable" apparently because of the lack of random sampling or control for pre-treatment differences, use of grade level equivalent scores, and the lack of a Finnish-speaking control group. It must also be pointed out that the Swedish speaking comparison students in this study scored well below national norms, but the results are certainly suggestive.

Croft and Franco (1983) reported that Spanish speaking children in a bilingual education program in New Mexico made better gains on the CTBS than a comparison group in grades 1,2 and 3, gaining an average of .3 more than comparisons over 7 months. Bilingual education students also made significantly better than "expected" gains in grades 4, 5 and 6. Randomization was not used, and grade level equivalent scores were reported, both violations of Rossell and Baker's criteria.

Medrano (1983, 1986) was not cited in Rossell and Baker. Medrano's subjects were 278 Mexican-American children taught in bilingual and non-bilingual programs. Medrano reported that the bilingual group was slightly, but not significantly better in reading, and significantly better in math at grade 3 (Medrano, 1983) and grade 6 (Medrano, 1986), controlling for grade 1 CTBS scores.

Fulton-Scott and Calvin (1983) compared students in bilingual education with two forms of English-only: pull-out ESL and "integrated ESL" (with native speakers of English). In a cross-sectional design, Fulton-Scott and Calvin examined grades and total CTBS examination results in grades one and six. There were no significant differences for any of the measures in grade 1, but bilingual education students earned significantly higher grades in grade six and their CTBS scores in grade six were significantly higher than children in the pull-out class. Effect sizes for grade six (table A.4) are very large.

Table A.4.  
ESL vs. Bilingual Education

grade	pull-out		ESL integrated		ESL bilingual education	
	mean	sd	mean	sd	mean	sd
1	1.56	1.14	1.06	.39	1.46	.35
6	4.89	.77	6.02	.79	6.85	.73

Effect sizes:

bilingual vs. pull-out ESL = 2.05

bilingual vs. integrated = 1.04

This study used grade equivalent scores, and would thus not be considered acceptable by Rossell and Baker.

Mortensen (1984) is listed in Rossell and Baker as "methodologically unacceptable," apparently because subjects were not randomly selected and no pretest was used as a covariate (Rossell and Baker cite this study as Mortensen's 1980 dissertation, not as a published paper). Mortensen compared

grade 4, 5, and 6 Spanish speaking students in two programs, a bilingual program with transitioning to English reading in grade 3, and a monolingual English program.

From the description provided, the bilingual program appeared to contain the three components considered to be characteristic of good bilingual programs (Krashen and Biber, 1988): literacy development in Spanish, instruction in Spanish in academic areas, and ESL.

Mortensen reported no difference between the groups on a "word attack" test, but the bilingual education students were significantly better on a test of comprehension skills (table A.5).

Table A.5.  
Performance on Word Attack and Comprehension Tests in English

	Number of Word attack skills mastered			Comprehension skills mastered	
	<i>n</i>	<i>mean</i>	<i>sd</i>	<i>mean</i>	<i>sd</i>
Bilingual	65	19.6	6.2	8.6	5.4
English-only	55	18.1	9.5	4.3	4.0

word attack:  $t = .94$ , *ns*

comprehension:  $t = 4.79$ ,  $df = 105$ ; effect size = .894

from: Mortensen, 1984

de la Garza and Medina (1985) is listed as an "acceptable" study in Rossell and Baker's bibliography but is not included in their analysis (they list it incorrectly as de la Garza and Marcella). de la Garza and Medina compared children in bilingual education to English-dominant children in an all-English program. Eighty percent of the bilingual education children were classified as "limited English proficient." The results were quite spectacular (table A.6).

Table A.6.  
Results of English Language Testing

<i>n</i>	Vocabulary		Reading Comprehension	
	<i>Bil. Ed.</i>	<i>English</i>	<i>Bil. Ed.</i>	<i>English</i>
	24	118	25	117
Grade 1	47.74	49.89	50.29	49.58
Grade 2	55.23	51.40	52.90	51.29
Grade 3	52.53	50.57	51.89	50.35

from: de la Garza and Medina, 1985

The children in the bilingual program scored as well as the English speaking comparison students and even outperformed them in the second grade vocabulary test. In addition, the SES of the English speaking children may have been higher (37% free lunches versus 76%; Medina and de la Garza, 1989).

There are flaws in the study: The controls in this study were not limited English proficient but were categorized as English-speaking. Students in the bilingual classes "volunteered to participate" (p. 116), and only 25 subjects of the original 76 had test scores available for all three years (24 for vocabulary). Nevertheless, the results are amazing and cannot be ignored.

So (1987) was a secondary multiple regression analysis of questionnaire data, from the High School and Beyond data base. So analyzed questionnaire results of students who had Spanish as their mother tongue and who had to take ESL classes. Students were divided into three categories based on their report of their education in grades 1 through 6: those whose education was all or almost all in English, all or almost all in Spanish, or "evenly mixed English/Spanish." Regardless of SES, those in the "evenly mixed" classes did better on tests of reading achievement. For low SES, all Spanish was better than all English, but for other levels of SES all English was better than all Spanish. Once again, however, mixed Spanish/English was better than both of the other treatments.

Few bilingual programs in the US are conducted entirely in Spanish. Thus, a plausible reason for the lower performance of the all-Spanish group is that this group probably represents a large number of recent immigrants who simply have not had sufficient time to acquire a great deal of academic English.

Krashen and Biber (1988), a report on bilingual programs in several school districts in California (Baldwin Park, San Jose, Fremont, Rockwood, San Diego), an individual school (Eastman), and a pre-school program (Carpinteria), clearly fails Rossell and Baker's criteria because random assignment was not used, nor were possible pre-existing differences in student achievement measured and controlled. In addition, statistical tests were not used. Once again, however, the results are solidly in favor of bilingual education.<sup>3</sup>

Gonzales (1989) used a design similar to the one used by Ferris and Politzer (1981). (Rossell and Baker cite Gonzales' dissertation, not the published report cited here.) He compared test performance of 34 sixth graders who had had at least two years schooling in Mexico and 38 sixth grade Spanish-speaking students who were born in Mexico but who had had all of their schooling in the United States. Both groups were enrolled in a bilingual program, but Gonzales' "Mexico" group had more literacy and subject matter instruction in Spanish. Table A.7 presents scores on a test of English literacy and an oral test of English.

Table A.7.  
Performance on Tests of English

test	Mexico		USA	
	mean	sd	mean	sd
English reading	60.32	12.70	53.05	17.04
English conv.	5.03	1.77	5.21	1.21

effect size for English reading: .48

Mexico: at least two years schooling in Mexico

USA: all schooling in USA

from: Gonzales, 1989.

The children who had had two years study in Mexico were slightly behind in English conversation, but both groups did very well: A perfect score on the test (the Bilingual Syntax Measure) is six. This confirms previous research showing that these children typically develop high levels conversational competence in English.

The Mexican group did significantly better than the USA group in English reading (Stanford Achievement Test), as well as on a test of Spanish reading. In addition, as others have found (see chapter 3), those who read better in Spanish also read better in English ( $r = .55$ ). Thus, more instruction in the primary language did not hurt: It helped.

An earlier version of Burnham-Massey and Piña (1990) was included as part of Krashen and Biber (1988). It is a report of bilingual education in the Baldwin Park Unified School District in California. The program in Baldwin Park comes close to the characteristics of optimal bilingual programs, as described in Krashen and Biber. Children in the bilingual program scored about as well as native speakers of English in their district on the CTBS (table A.8).

Table A.8.  
Long Term Results: Baldwin Park

test	group	grade 7	grade 8
CTBS Reading	Bilingual	35	38
	English	35	41
CTBS Language	Bilingual	56	46
	English	45	47
CTBS Math	Bilingual	61	63
	English	57	53

from: Burnham-Massey and Pina (1990)



Students in grades 7 and 8 had slightly higher grade point averages than the comparison students and outscored them on local tests of English reading, writing and mechanics. High school grades were equally impressive.

This study has the usual flaws: there was no comparison to limited English proficient children in an all-English program; comparison students were native speakers or English-dominant bilinguals. Also, the sample sizes in grades 7 and 8 were smaller than those reported in an analysis of the same cohorts in grade 5 (44 in grade 8, compared to 115 in grade 5 for the bilingual group), suggesting a selection bias. The results clearly show, however, that graduates of bilingual programs can do well.

Verhoeven (1991), not included in Rossell and Baker, studied 138 "working class" second grade Turkish-speaking children acquiring Dutch in the Netherlands. Several groups were studied. The first group consisted of two treatments: In grade 1, the submersion group first had instruction only in Dutch, the second language, followed by some instruction in Turkish literacy "for some hours per week". The transitional literacy group had Turkish literacy instruction along with oral Dutch. One subgroup continued with literacy instruction in both languages, adding Dutch after two months, while the other had only Turkish literacy until grade 2.

The second group "also followed a two-year transitional literacy curriculum" (p. 67).

In table A.9, I present Verhoeven's results for reading comprehension tested in Dutch, the second language, at grade 2. Note that both transitional literacy groups outperformed the submersion children. Statistical analysis showed that the scores were not significantly different, but children from the transitional class were better. (The effect size for the advantage of group 1 over submersion was a modest .38. For group 2 it was a more substantial .79).

Table A.9.  
Results: Reading Comprehension in Dutch after Grade 2

group	n	mean	sd
L1 literacy- Group 1	25	13.44	3.6
L1 literacy- Group 2	38	15.21	4.2
Submersion	74	11.93	4.1

from: Verhoeven, 1991

In table A.10, I present Verhoeven's results for attitudes toward reading. Questions on this measure dealt with preference among school subjects,

frequency of library visits, and free-time activities. Children who were in the transitional literacy program (only group 1 was tested) had significantly better attitudes toward reading in both Dutch and Turkish. If better attitudes and more interest in reading mean more reading, and if more reading means more literacy, these are important results.

Table A.10.  
Attitudes Toward Reading

	<i>Dutch</i>		<i>Turkish</i>	
	<i>mean</i>	<i>sd</i>	<i>mean</i>	<i>sd</i>
L1 literacy	12.00	2.4	12.89	1.5
Submersion	10.91	3.5	11.66	2.7

from: Verhoeven, 1991

Consistent with studies done in other languages, Verhoeven also reported a significant correlation between reading ability in Turkish and Dutch: Those who read better in their first language also read better in their second language.

Table A.11 summarizes our results thus far, listing all published studies that violated the selection criterion. Ten studies are positive (bilingual education superior), one shows no difference, and none are negative (all studies in Krashen and Biber other than Burnham-Massey and Piña are counted as one study). If we add Medrano (1983, 1986) and Fulton-Scott and Calvin (1983), the scorecard changes to eleven positive, two no difference, and no negative.

### Comments on Experimental Design

How serious is the failure to use randomization or the failure to control for possible pre-existing differences? In my view, it is important to note this flaw, but there are reasons to hypothesize that it is not fatal.

First, we have no reason to suspect that there were important differences among the groups; Mortensen (1984) reports that the children in her study all lived fairly close to each other and were of a homogeneous socioeconomic background and Ferris and Politzer (1981) report that the socioeconomic status of their Mexican-educated group was actually lower than their all US-educated group.

**Table A.11.**  
**Published Studies that Failed to Randomize or Otherwise Control for**  
**Pre-Existing Differences in Subjects.**

<i>Study</i>	<i>Other Flaw(s)</i>	<i>Results</i>
Kaufman, 1968		No difference
Rock Point, 1976, 78	GE scores, Norm-ref.	Positive
Ferris & Politzer, 1981		No difference
Gale et. al., 1981	Possible confounds	Positive
Lofgren & Ouvinen-Begerstam, 1982		Positive
Croft & Franco, 1983	GE scores	Positive
Mortensen, 1984		Positive
de la Garza & Medina 1985	Compared to "English-speaking" children	Positive
So, 1987		Positive
Krashen & Biber, 1988	Other treatments Norm-ref.	Positive
Gonzales, 1989		Positive
Burnham-Massey & Piña, 1990	Compared to "English-speaking" children	Positive
Verhoeven, 1991		Positive

*GE = grade equivalent scores used*

*norm-ref. = students compared to native-speaker norms*

Second, one can argue that with a large number of studies, randomization is present. If we look at many studies with non-random assignment, and have no reason to believe that subjects in different treatments differ in relevant ways, it can be argued that randomization of subject assignment has, in fact, occurred, because of the large number of studies. In other words, many slightly flawed studies can be combined to arrive at a valid analysis. In situations where randomization or pretesting are not possible, the answer is to use a post-test only design and to replicate many times.

Third, there is no evidence of selection bias in studies in which pre-test scores are available, that is, there is no consistent tendency for children placed in bilingual education to have higher pre-test scores.

- In Gersten (1985), in one cohort first-grade children in bilingual education scored 2.6 on the LAS for oral English, while English-only first-graders scored 2.7. In a second cohort, the respective scores were 2.28 and 2.18.

- Legarreta (1979) compared kindergarten children in submersion, submersion plus ESL, concurrent bilingual, concurrent bilingual with ESL, and bilingual without concurrent translation. Groups were not randomly assigned. Although all children were "identified as essentially monolingual in Spanish" (p. 524), the children in the concurrent translation plus ESL were lower in oral English comprehension at the start of the study. There were no differences among the other four groups.

- In Curiel, Rosenthal, and Richek (1986), there were no differences between students in bilingual education and English-only with respect to parents' education and books in the home. Students in the bilingual program spoke less English in the home.

- In Fulton-Scott and Calvin (1983) there was no significant difference among ESL and bilingual education children in CTBS scores in grade one.

In these studies, at least, there was no bias in favor of bilingual education.

### *Other Flaws*

In the first two sections of this report, I point out that some of the studies classified as "acceptable" by Rossell and Baker had serious problems. While they satisfied the requirements Rossell and Baker list, they had other problems. Do the "unacceptable" studies have these problems?

**Sample size:** I faulted Gersten (1985) because it compared 28 immersion children to 16 bilingual education students. The studies here contain more subjects. In addition, Gersten's measure (number of students who scored at or above grade level) makes the sample size problem severe, because a slight change in the scores of just a few students could have changed the overall results drastically.

**Duration:** Bilingual education often does not show its effects in early grades. Those studies in the "unacceptable" group that examined long-term

achievement (Gale et. al., Mortensen, Gonzales, the Rock Point study, Burnham-Massey and Piña) show positive effects of bilingual education while in the studies with shorter treatments the effects seem to be smaller (Verhoeven, Kaufman, Medrano, 1983). It should be noted, however, that de la Garza and Medina found strong effects in very early stages, Ferris and Politzer's study of junior high school students showed no difference, and Medrano's 1986 follow-up study found no difference for English reading (but a significant difference for math, favoring bilingual education).

**Control for socio-economic factors:** SES was not explicitly controlled in all the studies discussed here, but there is no reason to believe groups differed remarkably in this aspect. In Gersten (1985), experimental and control subjects came from different schools and might have had different amounts of English spoken in their environment.

**Lack of information about the bilingual program:** Several authors were explicit about the bilingual programs. Some appeared to be at least fairly consistent with current views on optimal programs (Gale et. al., Rock Point, Mortensen, Krashen and Biber) while others were not (Kaufman, Lofgren and Ouvinen-Birgerstam). The former group reported better results, but these were also the longer-term programs. Gersten (1985) tells us nothing about "bilingual education" in his study.

### **Irrelevant Studies**

A number of studies were categorized as "unacceptable" by Rossell and Baker, but were actually irrelevant to the purpose of their analysis. Their analysis was intended to focus on the effectiveness of TBE (transitional bilingual education) as shown by program evaluations. But some of the studies listed had very different goals:

Parr, Baca, and Dixon (1981) compared individualized and group instruction in a bilingual education setting. It did not compare bilingual education to non-bilingual approaches or to approaches utilizing less of the primary language. (They reported no difference between the two treatments.)

Chan (1981) is a comparison of Chinese-medium middle schools and bilingual (English-Chinese) schools in Hong Kong. No test scores for English language proficiency were included.

Ramirez and Politzer (1975) is not a program evaluation, but is an analysis of language use and language proficiency in Spanish and English. Their results are interesting: They found that home use of Spanish among elementary school

students was unrelated to English proficiency but was related to Spanish proficiency, suggesting that use of Spanish at home was not harmful to English. Use of English at home, on the other hand, resulted in poorer Spanish but was of no value to English proficiency.

Collier (1987) did not compare bilingual education with another program. Collier investigated the effect of age of arrival and length of residence in the United States among immigrant children who were not in bilingual education programs and concluded that it took from four to eight years to reach average levels on academic tests.

Escamilla and Medina (1993) reported on the impact of a bilingual education program on "limited" language proficient children (some oral ability in either English or Spanish) and "most-limited" language proficient children (low oral ability in both English and Spanish). It was not a comparison of bilingual education with other options. It did, however, address the issue of what to do with children who seem to lack competence in both languages. The good news is that both groups gained in oral competence in both languages. 94.5% of the most-limited students and 85.8% of the limited students gained in English oral ability over a three year span (K-2).

### Truly Unacceptable Studies

In some cases, Rossell and Baker are, in my view, correct in categorizing studies as "unacceptable," which means we learn nothing from them.

Golub (1978) claimed to be an "evaluation design" but had no control group and gave no scores on tests.

Trevino (1970) also had no control group and did not test for language.

Muller, Penner, Blowers, Jones, and Mosychuk (1977) is a comparison of children who participated in a Ukrainian-English (50-50) bilingual program and comparisons who were randomly selected from a group of students with similar socioeconomic backgrounds who were not in the program. Muller et. al. found no differences between the groups at the end of grade one in English language development. This study was probably classified as "unacceptable" because participation in the experimental group was not determined randomly nor was there any control for pre-existing differences. I would also classify this study as unacceptable: It is not clear whether it is a study of bilingual education or heritage language development. Only eight of the 20 students in the bilingual program came from homes in which both parents used Ukrainian in speaking to the child, and it is not clear, even in these cases, what level of competence the child had in Ukrainian.

Balinsky and Peng (1974) had no control group of children not in bilingual education, and used a translation approach (each lesson taught twice), which explains why their childrens' gains were low: for second graders, about 1-2 months gained in English reading in six months time.

### Conclusions on "Unacceptable" Research

Many of the published studies listed by Rossell and Baker as "unacceptable" are, to be sure, somewhat flawed, but a great deal can be learned from them. Despite the flaws, they are useful, and their consistent results cannot be ignored. In fact, the only published study in the unacceptable list that was negative was one I would also classify as unacceptable: Balinsky and Peng (1974), discussed just above.

A few studies were irrelevant. Their inclusion in the list of unacceptable studies gives the reader the impression that there are more unacceptable studies than there really are, and that the literature is of poorer quality than it really is. Interestingly, two of these studies provide evidence that supports bilingual education in other ways.

### The Final Score

Combining the acceptable and unacceptable studies discussed here, my final tally is as follows: 12 studies support bilingual education, 4 show no difference, and 2 are negative. Both negative studies are short-term (Moore and Parr; Gersten) and in both cases no description is given of the bilingual program. Of the four studies showing no difference, bilingual education is not described in any detail in two (Medrano, Rossell).

We should also consider the fact that method comparisons are not the only evidence we have supporting bilingual education. We have independent evidence that the principles underlying bilingual education are correct: There is strong evidence that background knowledge makes input more comprehensible (see Krashen, 1985 for a review) and that literacy transfers across languages (chapter 3, this volume). In addition, as argued elsewhere (chapter 2, this volume), the hypothesized principles underlying effective bilingual education predict quite accurately why some people appear to be successful without bilingual education. This is powerful triangulation.

### Notes

1. Rossell's report merits detailed discussion because of the importance attached to its conclusion in the popular press and the role it played in a court

decision (Teresa P. et al v. Berkeley Unified School District; US District Court, Case no. C-37-2346, DLJ).

To see how participation in a bilingual program affected performance, tables N.1, N.2 and N.3 present the regression coefficients computed by Rossell for participation in bilingual education. In all cases, relevant confounds such as

Table N.1.  
Regression Coefficients in Rossell (1990) for Participation in Bilingual Education

Gains in test scores Test Years		All Subjects, includes Chance Scores				Actual Number Tested			
		BE <sub>d</sub>	se	t	n	BE <sub>d</sub>	se	t	n
IPT, K-12	86-87	.522	.338	1.54	326				
IPT, K-6	86-87	.195	.385	.507	250				
CTBS reading	86-87	-2.83	6.63	-.427	111	1.90	6.10	.311	85
CTBS language	86-87	-1.65	6.52	-.253	112	6.92	6.40	1.08	85
CTBS math	86-87	3.85	8.49	.454	120	5.28	7.67	.688	89
CTBS reading	87-88	-11.95	4.98	-2.40	207	.626	4.42	.140	152
CTBS language	87-88	-13.24	5.37	-2.47	177	-3.63	4.99	-.727	152
CTBS math	87-88	-14.70	6.15	-2.39	208	-5.88	5.17	-1.14	154

BE<sub>d</sub>: regression coefficient for participation in bilingual education.

For sample sizes this large,  $t = 1.65$  required for .05 level, one tail;  $t = 1.29$  for .10 level.

socio-economic status (reflected by father's occupation) and age of students are controlled. The IPT is the IDEA Proficiency Test, which is given in the fall to all students considered potentially limited English proficient (Rossell, p. 86) in grades K through 12. The CTBS is administered in the spring, grades K-8 to students who score above a certain level in the IPT. For CTBS gains in table N.1, Rossell provides two sets of regression coefficients. The second set is based on the actual number of students tested, and is included in her paper in an appendix. The first set is based on more subjects; Rossell entered chance scores on the CTBS for those students who, on the basis of their IPT scores, were not eligible to take the CTBS at the start of the time span studied. In other words, gains are based on chance scores for the first analysis and actual scores for the second analysis of the CTBS.

From table N.1, it is clear that most of the regression coefficients are not statistically significant. As Rossell notes, the coefficients are significantly negative for gains on CTBS scores using the larger sample, suggesting that those who participated in bilingual programs gained less than those who did



not. (Note, however, that there is evidence of a slight superiority for bilingual education for IPT scores for students in grades 1-12; the regression coefficient for participation in bilingual education is positive and reaches the .10 level for a one-tail test. Rossell would probably call for a two-tailed test here, however.) It is very interesting that these negative results are reduced and sometimes disappear when the smaller sample is used, however.

Table N.2.  
Performance of reclassified students

Measure	Years	BE <sub>d</sub>	se	t	n
CTBS reading	80-87	2.97	5.86	.508	301
CTBS language	80-87	5.81	7.32	.793	296
CTBS math	80-87	12.22	7.42	1.65	302
Grades, reading	80-87	1.30	.847	1.53	347
Grades, language	80-87	1.03	.864	1.19	289
Grades, math	80-87	.052	.849	.060	354

In another analysis (see table N.2), Rossell compared California Test of Basic Skills (CTBS) scores for bilingual and ESL pull-out students after "reclassification." Rossell concluded that these data showed no difference between the two groups. For each subtest of the CTBS, however, the regression coefficient for participation in bilingual education was positive and in the case of math, it reached the .05 level for a two-tailed test, which Rossell did not indicate, and for reading grades the t-value reached the .10 level of significance, one-tail.

Rossell also compared Berkeley LEP children's performance on California Assessment for Progress (CAP) tests to performance by LEP children in two districts considered to have "exemplary" bilingual programs, Fremont and San Jose (Krashen and Biber, 1988). Rossell reported no significant difference among the children in the three districts in reading, and reported that the Berkeley students excelled in math.

There are problems with this conclusion. First, this analysis does not compare gain scores nor does it show how rapidly children reach norms. It considers LEP children as a group. The comparison is only valid if, in fact, LEP children in all three districts entered their respective systems at the same level of competence, and if all three districts used similar criteria for exiting children. This may not be the case. According to Rossell's analysis of reclassified children in Berkeley, many children scored very well on the CTBS long before they were exited - in CTBS Reading, for example, children in ESL pull-out

scored at the 33rd percentile two years before reclassification and at the 54th percentile one year before, while children in bilingual education who were reclassified scored at the 35th percentile two years before reclassification and near the 60th percentile one year before. CTBS Language Scores are similar, and scores in CTBS Math are even higher, with LEP children in Berkeley scoring above the 50th percentile four years before reclassification. Thus, Berkeley scores may look higher because some high-scoring children were retained in these programs longer.

Table N.3.  
Comparison of 3 Districts

Measure	Years	BE <sub>d</sub>	se	t	n
Reading CAP	85-87	11.82	11.58	1.02	108
Math CAP	85-87	2.357	12.335	.191	108

*n* = number of schools  
from: Rossell, 1990, tables 4.18, 4.19

Even if the analysis were a valid one - if children in all three districts entered at the same level and all three districts had equal reclassification criteria, it is interesting that, according to Rossell's analysis, schools that had bilingual education reported slightly higher CAP scores. Rossell's regression coefficients for bilingual education are presented in table N.3. The regression coefficients for bilingual education were positive (but did not reach statistical significance).

We are thus left with this picture of Berkeley: We have no idea how bilingual education was done in this district. According to one measure, Berkeley bilingual education students do not do as well as non-bilingual education students, but according to another (reduced sample) they do about as well (better in some measures, worse in others). Reclassification data gives bilingual students an edge, as does a comparison of LEP students across districts. This data hardly provides strong counterevidence to bilingual education.

2. The transitional vs. maintenance bilingual program study cited in Rossell and Baker is Medina and Escamilla (1992). TBE (transitional bilingual education) students (*n* = 125) were Vietnamese speakers; by grade 2, only 25% of their program was in the primary language. Maintenance students were Spanish-speaking (*n* = 298) and in grade 2, 60% of their program was in Spanish. There was evidence suggesting that the transitional students were of slightly higher SES: 55% of the transitional students received free or reduced price lunches, while 76% of the students attending the schools the maintenance students attended received free or reduced price lunches. Rossell and Baker claim that this study shows transitional bilingual education to be superior to

maintenance bilingual education in reading. This is not what happened. First, the measure used was oral language proficiency, not reading. Second, TBE was not superior. For the most limited students, there were no differences in gains in oral English. For the "nearly fluent," maintenance students were better both in kindergarten and grade 2. Superior gains were seen for the TBE students only in the most fluent group. Medina and Escamilla conclude that "results were mixed." (p. 282). In addition, the TBE students showed clear losses in their primary language, while the maintenance students did not.

3. Our report has been criticized by others as well. Chavez (1991) notes only that "the study of which (Krashen) is coauthor and which purports to demonstrate the practical effectiveness of keeping Hispanic children in extended bilingual programs is highly flawed. Moreover, it was published by the California Association for Bilingual Education, hardly a disinterested party in the debate" (p. 175). Chavez does not tell us what about the report is "highly flawed."

Imhoff (1990) maintains that the programs in Krashen and Biber (1988) worked because they were in "exemplary schools that are well-funded, staffed by highly trained and dedicated teachers, and composed of small classes of selected students" (p. 52). To my knowledge, not all of the schools described in our monograph were well-funded. The teachers did receive some extra in-service training in current theory and methodology, but to say they were more dedicated is not only unfounded but is also an insult to the teachers in the comparison groups. Nearly all of the students in the programs were unselected; there is no reason to suspect they differed from students in the comparison groups, and there is no reason to suspect differences in class size.

Rossell (1990) has also criticized our report, pointing out that one of the districts we studied, Fremont, took other positive action in addition to bilingual education (preschool, extra English reading, more parental involvement). While this could mean that these additional efforts were responsible for the Fremont children's outstanding performance, it is certainly not counterevidence to the hypothesis that bilingual education is effective.

Samaniego and Eubank (1991) raise several issues:

(1) Our analysis of Rockwood lacked controls. Thus, reported gains may not have been due to the treatment. But Rockwood students were compared to district norms (see our table 14). While district norms are not, strictly speaking, a control group, these comparison students had very similar backgrounds.

(2) They also were suspicious of a remarkable and "implausible" improvement made by Rockwood students (from the 6th to 38th percentile) from grades 3 to

6 in the 1981 cohort. Samaniego and Eubank claim that this result is even more remarkable because the sixth grade "had not even completed its first year of participation in the case study project" (p. 10). The project, however, began in Rockwood in 1981-82 and the sixth graders were tested in 1984.

(3) Their own analysis of data from Eastman and Rockwood led them to "strikingly different conclusions" (p. 10). Samaniego and Eubank compared sixth graders in Eastman in 1982, products of the old bilingual program, with 1986 sixth graders (who had received three years of instruction under the new plan) and found the median reading scores to be significantly different, in agreement with Krashen and Biber. In a similar analysis of Rockwood, however, Samaniego and Eubank reported that the 1982 sixth graders were significantly better in reading than the 1986 sixth graders. The data reported in Samaniego and Eubank's monograph for Rockwood differs, however, from the data made available to us. For example, in Samaniego and Eubank's table 4, the 1986 sixth graders achieved a median reading score at the 25th percentile on CTBS Reading, but according to our data their mean score was the 39th percentile. (We were not provided with 1982 sixth grade scores.)

Samaniego and Eubank present regression analyses of the Eastman and Rockwood data that provide strong support for the hypothesis that reading ability transfers across languages. Performance on reading tests in Spanish was a significant predictor of sixth grade English reading in all analyses; in the Eastman school, Spanish reading alone was a significant predictor, while at Rockwood, it was significant in interaction with mathematics ability, tested in Spanish. Samaniego and Eubank did not investigate transfer of mathematics performance from the first to the second language. Their explanation is interesting: "... there is considerably less doubt about the ability to transfer technical knowledge ... There can be no doubt that bilingual education makes a strong and important contribution to the ultimate development of math skills in LEP students ... The evidence which shows this is so overwhelming that it seems to us unnecessary to provide an analysis (similar to the one done for language) ..." (pp. 42-43).

## References

- Baker, K. and de Kanter, A. 1983. Federal policy and the effectiveness of bilingual education. In K. Baker and A. de Kanter (Eds.) *Bilingual Education*. Lexington, MA: DC Heath. pp. 33-85.
- Barik, H. and Swain, M. 1978. Evaluation of a bilingual education program in Canada: The Elgin study through grade six. Paper presented at the Colloquium of the Swiss Interuniversity Commission for Applied Linguistics. (ED 174 073).

- Barik, H., Swain, M., and Nwanunobi, E. 1977. English-French bilingual education: The Elgin study through grade five. *Canadian Modern Language Review* 33: 459-475.
- Balinsky, W. and Peng, S. 1974. An evaluation of bilingual education for Spanish-speaking children. *Urban Education* 9: 271-278.
- Bruck, M., Lambert, W., and Tucker, G.R. 1977. Cognitive consequences of bilingual schooling: The St. Lambert Project through grade six. *Psycholinguistics* 6: 13-33.
- Buriel, R. and Cardoza, D. 1988. Sociocultural correlates of achievement among three generations of Mexican American high school seniors. *American Educational Research Journal* 25:177-192.
- Burnham-Massey, L. and Piña, M. 1990. Effects of bilingual instruction on English academic achievement of LEP students. *Reading Improvement* 27:129-132.
- Chan, J. 1981. A crossroads in language instruction. *Journal of Reading* 24: 411-415.
- Chavez, L. 1991. *Out of the Barrio: Toward a New Politics of Hispanic Assimilation*. New York: Basic Books.
- Croft, D. and Franco, J. 1983. Effects of a bilingual education program on academic achievement and self-concept. *Perceptual and Motor Skills* 57: 583-586.
- Cummins, J. 1977. Delaying native language reading instruction in immersion programs: A cautionary note. *Canadian Modern Language Review* 34: 46-49.
- Cummins, J. 1981. The role of primary language development in promoting educational success for language minority students. In *Schooling and language minority students*. Sacramento, CA: California Department of Education. pp. 3-49.
- Cummins, J., Swain, M., Nakajima, K., Handscombe, J. Green, D., and Tran, C. 1984. Linguistic interdependence among Japanese and Vietnamese immigrant students. In C. Rivera (Ed.) *Communicative Competence Approaches to Language Proficiency Assessment: Research and Application*. Clevedon, England: Multilingual Matters.
- Curiel, H., Stenning, W., and Cooper-Stenning, P. 1980. Achieved reading level, self-esteem, and grades as related to length of exposure to bilingual education. *Hispanic Journal of Behavioral Sciences* 2: 389-400.
- Curiel, H., Rosenthal, J., and Richek, H. 1986. Impacts of bilingual education on secondary school grades, attendance, retentions, and drop-out. *Hispanic Journal of Behavioral Sciences* 8: 357-367.
- de la Garza, J. and Medina, M. 1985. Academic achievement as influenced by bilingual instruction for Spanish-dominant Mexican American children. *Hispanic Journal of Behavioral Sciences* 7: 247-259.
- Eagon, R. and Cashion, M. 1988. Second year report on a longitudinal study of spontaneous reading in English by students in early French immersion programs. *Canadian Modern Language Review* 44: 523-526.

El Paso Independent School District. 1987. Interim report of the five-year bilingual-education pilot. 1986-1987 School Year. El Paso, Texas: El Paso Independent School District, Office for Research and Evaluation.

El Paso Independent School District. 1989a. Bilingual education evaluation: The fourth year in a longitudinal study. 1987-1988 School Year. El Paso, Texas: El Paso Independent School District, Office for Research and Evaluation.

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

Escamilla, K. and Medina, M. 1993. English and Spanish acquisition by limited-language- proficient Mexican Americans in a three-year maintenance bilingual program. *Hispanic Journal of Behavioral Sciences* 15:108-120.

Ferris, M.R. and Politzer, R. 1981. Effects of early and delayed second language acquisition: English composition skills of Spanish-speaking junior high school students. *TESOL Quarterly* 15: 263-274.

Fulton-Scott, M. and Calvin, A. 1983. Bilingual multi-cultural education vs. integrated and non- integrated ESL instruction. *NABE Journal* 7:1-13.

Gale, K., McClay, D., Christie, M., and Harris, S. 1981. Academic achievement in the Milington bilingual education program. *TESOL Quarterly* 15: 297-314.

Genesee, F. 1983. Bilingual education of majority-language children: The immersion experiments in review. *Applied Psycholinguistics* 4:1-46.

Genesee, F., Tucker, G.R., and Lambert, W. 1978. An experiment in bilingual education: Report 3. *The Canadian Modern Language Review* 34: 621-643.

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

Gersten, R. and Woodward, J. 1985. A case for structured immersion. *Educational Leadership* 43: 75-79.

Gersten, R., Woodward, J., and Schneider, S. 1992. Bilingual Immersion: A Longitudinal Evaluation of the El Paso Program. Washington, DC: The Read Institute.

Gonzales, L.A. 1989. Native language education: The key to English literacy skills. In D. Bixler-Marquez, J. Ornstein-Galicia, and G. Green (eds.) *Mexican-American Spanish in its Societal and Cultural Contexts*. Rio Grade Series in Languages and Linguistics 3. Brownsville, Texas: University of Texas - Pan American. pp. 209-224.

Golub, L. 1978. Evaluation design and implementation of a bilingual education program, grades 1-12, Spanish/English. *Education and Urban Society* 10: 363-383.

Holobrow, N. Genesee, F., Lambert, W., Gastright, J. and Met, M. 1987. Effectiveness of partial French immersion for children from different social class and ethnic backgrounds. *Applied Psycholinguistics* 8: 137-152.

Hoover, W. 1983. Language and literacy learning in bilingual instruction. Austin: Southwest Educational Development Laboratory.

Imhoff, G. 1990. The position of U.S. English on bilingual education. In C. Cazden and C. Snow (eds.) *English Plus: Issues in Bilingual Education*. Annals of the American Academy of Political Science. Newbury Park, CA: Sage Publications. pp. 48-61.

Kaufman, M. 1968. Will instruction in reading Spanish affect ability in reading English? *Journal of Reading* 11: 521-527.

Krashen, S. 1985. *The Input Hypothesis*. Beverly Hills, CA: Laredo Publishing Company.

Krashen, S. 1993. *The Power of Reading*. Englewood, CO: Libraries Unlimited.

Krashen, S. 1994. The input hypothesis and its rivals. In N. Ellis (ed.) *Implicit and Explicit Learning of Languages*. London: Academic Press. pp. 45-77.

Krashen, S. and Biber, D. 1988. *On Course: Bilingual Education's Success in California*. Ontario, CA: California Association for Bilingual Education.

Legaretta, D. 1979. The effects of program models on language acquisition by Spanish speaking children. *TESOL Quarterly* 8: 521-576.

Lofgren, H. and Ouveinen-Birgerstam, P. 1982. A bilingual model for the teaching of immigrant children. *Journal of Multilingual and Multicultural Development*. 3: 323-331.

Medina, M. and de la Garza, J. 1989. Initial language proficiency and bilingual reading achievement in a transitional bilingual education program. *NABE Journal* 13: 113-125.

Medina, M. and Escamilla, K. 1992. Evaluation of transitional and maintenance bilingual programs. *Urban Education* 27: 263-290.

Medrano, M. 1983. The effects of bilingual education on reading and mathematics achievement: A longitudinal case study. *Equity and Excellence* 21: 17-19.

Medrano, M. 1986. Evaluating the long-term effects of a bilingual education program: A study of Mexican American students. *Journal of Educational Equity and Leadership* 6(2): 129-138.

Mezynski, K. 1983. Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research* 53: 253-279.

Moore, F., and Parr, G.. 1978. Models of bilingual education: Comparisons of effectiveness. *The Elementary School Journal* 79: 93-97.

Mortensen, E. 1984. Reading achievement of native Spanish-speaking elementary students in bilingual vs. monolingual programs. *The Bilingual Review* 11: 31-36.

Muller, L., Penner, W., Blowers, T., Jones, J. and Mosychuk, H. 1977. Evaluation of a bilingual (English-Ukrainian) program. *Canadian Modern Language Review* 33: 476-485.

- Parr, G., Baca, F., and Dixon, P. 1981. Individualized versus group instruction in bilingual education: A two year study. *The Elementary School Journal* 81: 223-227.
- Porter, R. 1990. *Forked Tongue: The Politics of Bilingual Education*. New York: Basic Books.
- Ramirez, A. and Politzer, R. 1975. The acquisition of English and the maintenance of Spanish in a bilingual education program. *TESOL Quarterly* 9: 113-124.
- Rosier, P. and Farella, M. 1976. Bilingual education at Rock Point: Some early results. *TESOL Quarterly* 10: 379-388.
- Rossell, C. 1990. The effectiveness of educational alternatives for limited-English-proficient children. In G. Imhoff (ed.) *Learning in Two Languages*. New Brunswick, NJ: Transaction Publishers. pp. 71-121.
- Rossell, C. and Baker, K. 1996. The educational effectiveness of bilingual education. *Research in the Teaching of English* 30: 7-74.
- Schorr, B. 1983. Grade school project helps Hispanic pupils learn English quickly. *The Wall Street Journal*.
- Samaniego, F. and Eubank, L. 1991. A statistical analysis of California's case study project in bilingual education. Technical Report Series of the Intercollege Division of Statistics, University of California at Davis.
- Smith, F. 1994. *Understanding Reading* (fifth edition). Hillsdale, NJ: Erbaum.
- So, A. 1987. Bilingual education and hispanic reading achievement. *Contemporary Education* 59: 27-29.
- Trevino, B. 1970. Bilingual instruction in the primary grades. *Modern Language Journal* 54: 255- 256.
- Verhoeven, L. 1991. Acquisition of biliteracy. *AILA Review* 8: 61-74.
- Vorih, L. and Rosier, P. 1978. Rock Point Community School: An example of a Navajo-English bilingual elementary school program. *TESOL Quarterly* 12: 263-69.
- Willig, A. 1985. A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research* 55: 269-317.
- Willig, A. 1987. Reply to Baker. *Review of Educational Research* 57: 363-376.