

Ramírez et al.: Misled by Bad Theory

Keith Baker

Abstract

Two shaky theories dominate the debate over bilingual education. The facilitation hypothesis predicts a long-term advantage for bilingual education over all-English instruction. The time-on-task hypothesis predicts that all-English instruction is superior to bilingual education in teaching English. The data contradict both theories. Bilingual education programs are superior to all-English instruction in the early stages of learning English.

Introduction

Although not much discussed in the final report (Ramírez, Pasta, Yuen, Ramey & Billings 1991) the design of the Ramírez study took bilingual education program theory strongly into account. The third grade cohort was included in the design to create a synthetic cohort estimate, by combining the late-exit kindergarten and third grade cohorts, to test Cummins' facilitation hypothesis [which is detailed later on (Cummins 1978, 1981, 1985)]. Learning curves over time were made a central part of the analysis to test both Cummins' theory and the alternative time-on-task theory that predicts immediate and continuing superiority for all-English programs. These theoretical considerations guided the analysis and the interpretation of the data.

This paper will show that the theoretical framework of the study -- a comparison of the facilitation hypothesis and the time-on-task hypothesis -- led Ramírez, Pasta, Yuen, Ramey & Billings to pay too much attention to a very weak, insupportable finding in the Trajectory Analysis of Matched Percentiles (TAMP) curves and to over look a major finding with strong empirical support which fit neither theory.

My article has three parts. First, the empirical findings of Ramírez et al. are discussed. This discussion shows their conclusion that there were long-term effects for the late-exit program is highly speculative and probably not valid. I also show that the Ramírez et al. data show that *bilingual education programs are superior to all English instruction in the early stages of learning*

English. Ramírez et al. pay little attention to this important finding, probably because it fits neither of the theoretical perspectives they were considering.

The second part of the article discusses the two theories these researchers were considering. I show that both theories lack empirical support, not only from the Ramírez study but also in the rest of the literature as well. Given how poorly supported the two theories are in the literature, the finding of short-term, immediate benefits for bilingual education programs merits far more attention than Ramírez et al. gave it.

In the third part of the article, I show that the immediate, short run superiority of bilingual education programs has a strong theoretical basis in experimental psychology and is also found in about a dozen other studies that have been generally ignored because they also did not fit the prevailing theoretical model.

Ramírez et al.: The Comparative test score analysis

Ramírez et al. found no effects for language or math or for reading over the first four years of school. However, from Kindergarten to grade one, the first two years of school, these researchers found a significant effect favoring bilingual education programs. Since the full report of this study is not easily available, the results of the hierarchical linear model analysis of the K-1 (spring) reading data from the California Test of Basic Skills (CTBS) are presented in Table 1. It is clear from Table 1 that bilingual education programs were superior to all English instruction over the first two years of schooling. It is equally clear from the analysis in Ramírez et al. that this advantage of bilingual instruction had vanished by the end of four years of schooling (see Table 2). Ramírez et al. pay little attention to initial superiority of bilingual education programs.

The Late-exit analysis

The late-exit programs were all located in districts with no other program operating. *The analysts therefore chose not to compare the late-exit program to the immersion or early-exit programs.* Considering the exaggerated claims for the late-exit program model that have since been made, this was unfortunate. I calculated the average gains on tests in English over the first four years of the program, which are shown in Table 3.

Table 1
Reading results at the end of grade one by analytic model

Students	Independent variables in the analysis	Parameter*	I-test	p=
All	Program only	7.147	1.852	.065
All	Program & school	8.977	2.367	.019
All	Key covariates	3.177	2.167	.031
All	All covariates	8.433	2.223	.027
Pretest	Program only	13.068	2.670	.008
Pretest	Program & school	14.935	2.861	.005
Pretest	Key covariates	14.619	2.794	.006
Pretest	All covariates	16.040	3.000	.003
Pretest	Program & both pretest	12.677	2.598	.010
Pretest	Program, school & both pretests	13.924	2.675	.008
Pretest	Key covariates & both pretests	13.707	2.623	.010
Pretest	All covariates & both pretests	15.552	2.920	.004
Pretest	Program & pretest total	12.360	2.537	.012
Pretest	Program, school & pre total	13.817	2.659	.009
Pretest	Key covariates & pre total	13.586	2.605	.010
Pretest	All covariates & pre total	15.259	2.874	.005

* The parameter may be read as the difference between the bilingual and all English program (+ values favor bilingual education), adjusted for the listed covariates. Source Ramírez et al., 1991, Table 88.

Table 2
Reading gains between grades one and three

Independent variables in the analysis	Parameter	t-test	p=
Program	5.165	1.196	.233
Program & school	6.585	1.552	.122
*Key covariates	4.784	1.130	.260
All covariates	5.459	1.289	.199

* The parameter may be read as the difference between the bilingual and all English program (+ values favor bilingual education), adjusted for the listed covariates. Source Ramírez et al., 1991, Table 88.

Table 3
Gains through the third grade for three programs

	Mathematics	Language	Reading
Immersion	44.0	72.9	63.7
Early -Exit	47.1	68.3	60.4
Late-Exit	33.3	43.9	51.4

Source: Adjusted tables from vol. 2, Appendix C.

It seems clear that late-exit students fell well behind early-exit and immersion students during the first four years of elementary

school in their growth in English. The study shows with statistical analysis that the difference between the immersion and early-exit programs was not statistically significant, that is, the differences that occurred probably are due to chance variation between the two samples. However, the study did no statistical comparison of the late-exit programs to either of the other two types. Nevertheless, the magnitude of the difference between the late-exit gains and those of the two other programs suggests that the difference is real.

The report points out that late-exit programs, unlike early-exit or immersion programs, have a positive curvature over time; that is, the rate of gain increases from year to year. However, the report's conclusion that this difference in curvature represents an advantage for late-exit programs does not necessarily follow. Late-exit students may be playing catch-up. They appear to have fallen well behind their immersion and early-exit peers through the third grade, and it is indeed fortunate for them that they do catch-up later on. *Unfortunately, the data available does not make it possible to determine if they ever do overcome the handicap they appear to have acquired during the first three years of the late-exit program.*

The differences among the programs were small when compared to the differences between schools, that is, it makes a much bigger difference *where* a limited English proficient (LEP) student goes to school than which program he is in.

Trajectory analysis of matched percentiles (TAMP) analysis

In addition to the statistical analyses discussed above, the report also contains an extensive graphic analysis. The graphic analysis has two purposes: (a) to confirm the statistical analysis by looking at the same test score data with a different analytic method and (b) to explore in more detail the relationships among the variables.

Unwarranted statements about the effects of late-exit programs have been made by misinterpretation of this analysis. Therefore, we will take a closer look at it. First, the cautions of the report on the use of this analysis are worth repeating: "These graphs should not be considered a replacement for the analyses of the earlier chapters.. The graphs in this chapter do not reflect any adjustment for differences among students, schools, and districts and are, therefore, potentially misleading" (Ramírez et al. 1991 p. 353). "The TAMP figures cannot answer whether any growth differences between any two districts are attributable to differences in students,

school, or district characteristics or are attributable to the bilingual program” (Ramírez et al. 1991 p. 367).

Findings that appear to support the effectiveness of the late-exit strategy are based on comparisons of districts within the late-exit sample, precisely the type of comparison the report warns is not justified.

Comparing programs for LEPs to the national norm is not proper, as the Office of Bilingual Education and Minority Language Affairs (OBEMLA) recognized a few years ago when OBEMLA recommended the Gap Reduction model over norm-based comparisons (Sec. 151 Model A) for evaluating bilingual education programs. An evaluation of the early bilingual education program in San Jose, California, illustrates the basic problem. The students in the bilingual education program gained 32 percentiles, which seems impressive until you learn there was also a comparison group of students who were also LEP language-minority children but who were not in any special program -- they were in a “sink or swim” environment. They gained 28 percentiles, not statistically different from the gain of the bilingual education program students. Both the LEP language-minority children in the bilingual education program and in “sink or swim” made equally big gains in their English. Since both groups made equal gains in English, their gains in English had to come from something other than the bilingual education program.

Norm-based comparisons should not be made in bilingual education program evaluations because they represent student growth, not program effects. Consider a non-English speaking student who was never in a special bilingual education program who learned English (since bilingual education programs were rarely if ever used for immigrants before the late 1960s, we know it is possible to learn English, and to learn it well without bilingual education programs). We will test this hypothetical student twice, once when he first enters the U.S. and a second time a few years later. As a new immigrant knowing no English, his first score is 0. After some time in the country, he learned English and scored at the 50th percentile. These scores track the student’s progress in learning English from somewhere. However, norms cannot differentiate the learning of English from TV, from friends on the street, from the playground at school, from no special school program, or from a bilingual education program.

Normative comparisons tell us the student is learning English, but not the effects of any particular program on learning English. The only way to determine if a bilingual education program is effective is to compare a group of LEP students in the program to another group of LEP students enrolled in no special program. Since the law requires LEPs to be in special programs, it is nearly impossible to determine if bilingual education programs are effective. The study design of Ramírez et. al. used the next best alternative, comparing the relative effectiveness of different programs to each other. If we assume that some program is better than no program, then the best program can be assumed to be effective.

To illustrate the problems encountered in looking at gains in relation to the norm, as was done in the report, consider the example in Table 4.

Table 4
Hypothetical example of normative gains

	Program Percentile	Program Gain	Norm Percentile	Norm Gain
First Test	22		50	
Second Test	24	2	50	0

The norm is fixed so that there is no gain from year to year in the norm; the average score is always the 50th percentile for the norming group. The hypothetical program in the table showed greater gains than the norm, but is it a good program? It is not a good program if we expect the students in the program to have an achievement level comparable to that of the norm because, despite the greater gain, the program's students scored well below the norm. Gains against the norm are useful in the proper context, but must be looked at carefully and with great trepidation.

For these reasons, the main analysis of the study focused on comparing programs and not on norm comparisons. For these reasons, the U.S. Department of Education's Request for Proposals (RFP) for the study cautioned the researchers not to use normative comparisons in the analysis unless they were supported with an extensive battery of other analyses which the report does not present. Norm comparisons can tell us only one thing about the effects of bilingual education programs: if the students in the program are making gains at a rate less than that of the norming

group, it is a signal that the program needs a closer look; something may be wrong. The reason for this is that students who do not speak English will almost always make greater gains than the norm group as they learn English from any and every source. The TAMP curves can check for disasters.

To make Table 5, I studied the TAMP graphs in the report and noted whether the program group gains were A(bove), (equal to), or B(elow) the norm group. In some cases the pattern changed at different scores so that some curves might show an initial pattern of gains above those of the norm followed by gains below the norm for others in the group. In this case, the Table would be marked A,B (gains Above the norm at first followed by gains Below the norm).

Table 5
Gains of Program students compared to the norm

Time Period (in grades)	Immersion	Early-Exit	Late-Exit
	Math	Math	Math
K-1	A	A	A
1-3	B	B	B
	Language	Language	Language
K-1	A	A,=	A
1-3	B, A	B, A	B
	Reading	Reading	Reading
K-1	NA	NA	NA
1-3	A		

Source: TAMP curves (A= above the norm; B= below the norm; equals the norm)

The graphic analysis indicates some areas of concern. There is a problem in math instruction in all programs. This suggests the source of the problem is not related to the language differences among the programs. For some reason, Hispanic LEP students start school behind the general population in math and that deficiency gets worse over the first three years of school regardless of the type of language program used.

All three programs perform acceptably in language during the first year of school (K-1), but then danger signals arise. For grades 1-3, none of the programs is consistently above the norm rate of growth and the late-exit programs are generally below the norm. In English reading, only the immersion program attained the minimum level of acceptance.

In light of what the data shows, serious questions must be raised about the widely reported conclusions of the U.S. Department of Education that their policy is a success because all programs are performing (showing gains) above the norm. The widely reported evidence of success for late-exit programs comes from only three of over 200 graphs used in the TAMP analysis.

The TAMP analysis is, as the report recognizes, highly speculative. Conclusions are drawn about relative program effectiveness entirely from data on just one of the three programs. The superficial superiority of the one pure case of late-exit programs is not supported by a closer look at the data. In short, the report's conclusions about late-exit programs are invalid.

In summary, I agree with Ramírez et al. on these conclusions:

- After four years of school, there are no program differences.
- There is no evidence supporting the time on task theory that all-English instruction is superior to bilingual education programs.
- There is no reason to prefer all-English programs over bilingual education programs.

I disagree with Ramírez on these points:

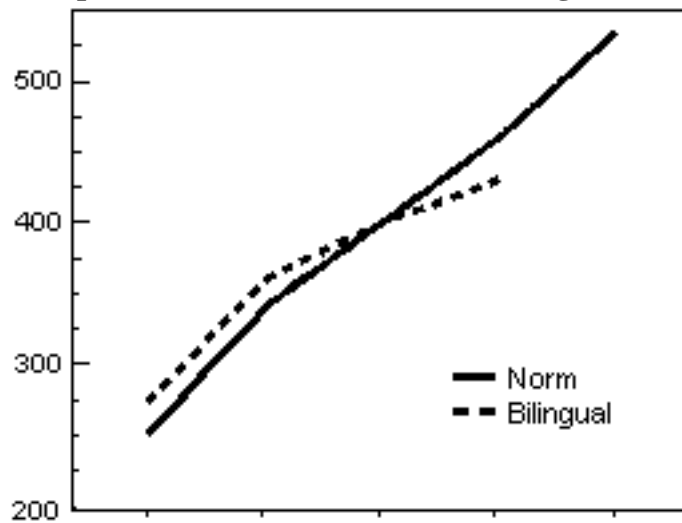
- There is no advantage for late-exit programs.
- The most important finding was the superiority of bilingual education programs over all-English instruction after two years of school.

Evidence of the effect of bilingual education programs in the early stages of learning English is also seen in Fig. 1 from Ramírez et. al.'s TAMP analysis. Fig. 1 shows that greater gains between the first and third grade were made by LEPs with the lowest English proficiency in grade one. Those LEPs with the highest proficiency in grade one showed gains below the normative rate through grade three, whereas the lowest proficiency LEPs had gains greater than the norm.

The rest of this article discusses the last point and argues that bilingual education programs are superior to all-English instruction in the early stages of learning English. Ramírez et al. overlooked the importance of this finding, not because they are champions of bilingual education programs as some in the English-only movement have suggested, but because the analysis was done within a theoretical framework which did not allow for such a finding. Since the finding did not fit theory, Ramírez et al. more or less dismissed

it as being unimportant. However, the rule in research is that if the facts do not fit the theory, it is the theory that is wrong. Therefore, the finding of early superiority for bilingual education programs over all-English instruction is even more important because it has major theoretical as well as instructional and policy implications.

Figure 1
Tamp Curves: Grade 3 vs. Grade 1 English Reading



Source: Ramirez et al. (1991), Fig. 100

Because there are implications of the finding that bilingual education programs are superior to all-English instruction early in the learning of English, it is worth taking the time to show that this is not an isolated finding. Indeed, the early superiority of bilingual education programs is well founded in the literature.

Additional evidence of the early benefits of bilingual education programs

Another recent study (Burkheimer, Conger, Dunteman, Elliott, & Mowbray 1990) will be described. Burkheimer et al. (1990) tested about 10,000 LEPs in 84 schools over three years. However, their analysis as reported, looked at only about 4,000 of those students; students who remained in the special language program for all three years of testing. The students were enrolled in a variety of programs, ranging from all English to extensive bilingual.

maintenance. Altogether, eight basic program models and five sub-variations were included in the study. The major dimensions along which the programs varied were the extent of use of the first language (L1) and the extent to which English instruction was structured for those learning English as a second language (ESL).

Burkheimer et al. found a large effect for use of L1 in the second grade for those students who initially scored low on English. There was no effect for the use of L1 in grades 3-6. Grade 1 was not tested. Their findings are reproduced in Table 6.

Table 6
Significant coefficients in Burkheimer et al.

	First Grade Cohort	First Grade Cohort	Third Grade Cohort	Third Grade Cohort	Third Grade Cohort
Post-test in study year:	2	3	1	2	3
Grade in school	2	3	3	4	5
Raven (non-verbal 10)	.150	.119	.078	.120	
Initial Oral English		.116	.102	.135	
Initial oral L1	.068*	-.843	.050*		
Age					
Sex					
% of life living in U.S.	-.054				.102
Family SES index					
Use of English at home					
Language re-test	.342	.437	.676	.506	.582
% LEPs in student body of school	-.164				
Neighborhood SES			.067		
Class size					
% of class that are LEP			-.442		
Total hrs. a week of oral English instruction			-.148		
Hrs. a week of math		-.42*			
Hrs. of math in English minus hrs. of math in L1	.303	.538	-.17*		
Hrs. of language (excluding oral English)	-.173	-.622	-.362		
Hrs. English lang. arts minus hrs. L1 lang. arts	-.497*	.110			.208
Ethnic heritage taught	.220	.231	-.151		.1
Simplified English pre-dominant in teaching					

Hrs. English lang. arts taught in L1	.506				
School & district dummy variables					
	-.195*				
	.2				
	.209				
	-.161				
				-.2*	
				.226	
	.256				
				-.184	
	-.215				
				.213*	
				.302*	
				.198*	
	-.105				
	-.198				
	-.077				
		-.270		.176	
		.587			
		.694			
Yrs. enrolled in the current school					-.134
Lang. arts teacher had special training for LEPs	-.085				
Lang. arts teachers familiarity with L1	.132				
Parents opinion of how student likes school					.084
Use of reading aloud, workbooks, & listening in oral English instruction	.106				
Use of translation or memorizing in oral English instruction	.100				

Burkheimer et al. conducted an iterative series of regression analyses to arrive at the final model shown in Table 6. The most striking feature of the Table is the lack of consistency in the effects of any variables across grades except for pre-test score and non-verbal intelligence.

We are interested in the variable “hours of English language arts that were taught using the native language in a week” ($b=.506$). Burkheimer et al. note that there was an interaction between this variable and the English oral pre-test and the English language arts test. Therefore, they conducted a special analysis to take the interactions into account by dichotomizing the distribution for the pre-test measures of English proficiency and found that use of L1 has a large positive effect on students who initially scored very low in English, but a large negative effect on students initially scoring high in English. Note that none of these effects occurred in later grades, only in the youngest students in the study.

Burkheimer et al. show that bilingual education programs are superior to all English instruction for certain LEPs. It seems these students can be characterized as in the early stages (the first 2-3 years) of learning English. That bilingual education programs positively effect the early stages of learning English is inferred from the Burkheimer findings and that the effect was limited to (1) the second grade for students who had been in programs since they entered school, e.g., since Kindergarten or first grade, and (2) LEP students who scored lowest in English at the pre-test or who scored highest in L 1. There was no effect for bilingual education programs for second graders who had relatively high scores in English at the pre-test or for any students in grades 3-6.

Further support for the early effectiveness of bilingual education programs early in the process of learning English is found in Barkley (1983) who tested 67 Hispanic students in a summer Head Start program with a language training emphasis. The students were randomly assigned to eight groups; four treatments for teaching English as a second language (L2) vocabulary development (bilingual teaching, all English, Spanish only to introduce and define the English vocabulary words, or a control treatment arts and crafts class) for each of two teachers. Treatment was for two short periods within a 35 minute class, then all the students returned to the regular class. The students were tested at the end of the summer session and eight months later at end of kindergarten. The analysis was a meta-analysis on 14 measures. Barkley concluded “teaching English vocabulary words by introducing them and explaining them in the children’s native Spanish leads to more English vocabulary.. .than any other method” for “young children who are native in Spanish and not yet fluent in English” (p. 365).

Again, there were positive effects for using L1 on learning English in early English learners and the results were immediate, they did not show up years later as predicted by the facilitation hypothesis. Again, Spanish was used in a way that produced strategically timed rest periods in the English learning task; spaced rather than massed learning.

Finally, Baker & de Kanter's (1981; 1983) review of 39 methodologically sound studies of the effects of bilingual education programs, found *11 studies reported short-term positive effects for bilingual education programs*. Positive effects (over one year or less of testing) were reported early in school as well. Effects in kindergarten and grade one were found by the American Institute for Research (AIR 1975); in kindergarten by Legareta (1979); in pre-kindergarten through the third grade by McConnell (1980); in the first grade by Morgan (1971); and in grades one through three by Zirkel (1972).

When we designed the Ramírez study, we paid special attention to theoretical models of bilingual education. The study was designed to test two competing theories of how LEPs learn English. The two competing theoretical positions are the facilitation hypothesis put forth by advocates of bilingual education programs and a time-on-task hypothesis put forth by advocates of all-English instruction for LEPs.

Because these theories were so important in the design of the study and strongly influenced the interpretation of the data, they will now be discussed in some detail.

Theoretical models of LEP learning

The Facilitation hypothesis

Ramírez et al. are not alone in dismissing the evidence of early effects for bilingual education programs. Cummins (1978) argued that bilingual education programs have only long-term effects. This view, although lacking empirical support, came to so dominate thinking on bilingual education programs that the field in general overlooked the immediate superiority of bilingual education programs.

With respect to literacy, advocates of bilingual education programs theorize that there is a facilitating effect of the first language on second language learning. It is argued that it is easier to learn how to read in a language you already know than in a new

language. Once a child has learned how to read in his native language, developing literacy in the second language is facilitated because the learner is already literate in one language. Thus, a Spanish speaking child who first becomes literate in Spanish and then learns English will, after some time, have better mastery of English than will a student taught in English from the outset.

It is hypothesized that developing literacy in a second language is faster than in the first language because many metalinguistic concepts and elements of literacy transfer quickly from the first language to the second. For example, the concept of sentence or paragraph need only be learned once. Having learned what a paragraph is in Spanish, it is obvious to a Spanish speaking learner of English that an English paragraph is little different from a Spanish paragraph in its meaning and use in reading and writing. The hard learning task--the time consuming task--is to get the child to grasp what a paragraph is for the first time.

An important feature of this proposition is the temporal pattern involved. Literacy must first be developed in L1, then in L2. Sometime in the process of learning L2 the facilitating effects will show up. Since it takes 3-4 years to initially acquire literacy, the facilitating effect will not become apparent until 5 years or so after literacy instruction begins.

James Cummins (1978, 1981, 1985) has been the principal exponent of the facilitation theory. Cummins' (1978) initial theoretical work addressed the problem of inconsistencies in the empirical research. He set out to offer an explanation for the conflicting findings of studies of the effects of bilingual education programs versus instruction all in the second language. Some studies found bilingual education programs superior, other studies found them inferior. Cummins argued there was a threshold level of L1 literacy that had to be attained before the facilitation would take effect to expedite learning a second language. Thus, the conflicting research was explained by whether or not the L1 threshold had been reached by the bilingual learners. Studies finding in favor of all-L2 instruction had subjects in the bilingual education program who had not reached the threshold in L1 while studies where bilingual education was more effective than all-L2 instruction had subjects in the bilingual education program who reached the threshold.

If bilingually taught children have reached the threshold in their native language, then they will learn the second language better than will students taught entirely in L2. On the other hand, if bilingually

taught children have not reached the threshold in L1, they will be inferior to students taught entirely in L2.

The principal evidence Cummins cited for this idea was Skutnabb-Kangas and Toukomaa (1976) who looked at students who had immigrated from Finland to Sweden. Two groups of immigrants were compared on their performance in Swedish (L2), those who immigrated before reaching grade three in school and those who immigrated during the third grade or later (mastery of L1 literacy). Students who immigrated later, that is, those students who had been in school in Finland long enough to have first developed literacy in their native language (Finish), supposedly performed better in Swedish than did the children who had moved to Sweden at a younger age and who presumably began learning Swedish at an earlier age.

There are major methodological problems with Skutnabb-Kangas and Toukomaa and with the inference that its results support the facilitation hypothesis. A detailed discussion of these problems is in Baker and de Kanter (1981) and in Baker (1984), only two are mentioned here. First, Skutnabb-Kangas and Toukomaa reported no statistical analysis of their data. Baker and de Kanter (1981) conducted an extensive statistical analysis of the Skutnabb-Kangas and Toukomaa data and found there was no relationship between length of schooling in the native land and later performance in L2.

Moreover, at the time of the Skutnabb-Kangas and Toukomaa study, Swedish, the second official language of Finland, was a required subject in Finish schools from the third grade onward. Thus, if the Skutnabb-Kangas and Toukomaa data show anything, what they show is that students who have a chance to study a second language before immigrating perform better in that language than do students who had no formal instruction in the language before they immigrated. In short, there is no empirical support in Skutnabb-Kangas and Toukomaa for the facilitation hypothesis.

More than a decade of research and literally thousands of studies, since Cummins first proposed his theory, have confirmed neither the theory nor the effectiveness of bilingual education programs in the long run. There have been a number of reviews and discussions of the effectiveness of bilingual programs on performance in English and other academic subjects (Troike 1978; Baker & de Kanter 1981, 1983a,b; Rossell & Ross 1986; Rotberg 1982; Willig 1985; Yates & Ortiz 1983; Peterson, Berry, Abbott, Kravant, Sundusky, Chow & Ortega 1976; Holland 1986; Ravitch

1983; Dulay & Burt 1978) and the following conclusions can be drawn from this literature:

- Poor study design and poor methodology abound.
- Bilingual educators and program advocates reach far more positive conclusions when reviewing the literature than do reviewers from outside the bilingual education field.
- Reviewers from outside the bilingual education field (Baker & de Kanter 1981, 1983; Rossell & Ross 1986; Rotberg 1982; Peterson et al. 1976; Ravitch 1983; Holland 1986) are quite pessimistic about the effectiveness of bilingual education.
- The most positive thing that can be said about bilingual education from these reviews is that its effectiveness in meeting the special needs of LEPs remains to be proven.
- Most bilingual programs have no effect on raising performance levels of English and other academic subjects. Some programs have a positive effect; some programs have a negative effect.

In spite of the lack of empirical support for Cummins' post-hoc theorizing, the facilitation supposition has been overwhelmingly accepted as fact by bilingual educators.

The Competing theory

Just as the advocates of bilingual education programs have a theory to support their beliefs, so to the opponents of bilingual education programs have a theory with which they attack bilingual education programs and the facilitation idea. Opponents of bilingual education programs counter the facilitation hypothesis with a time on task hypothesis, arguing that learning English is a function of time spent studying English (see Porter 1990, for an example). Therefore, bilingual education programs are inferior to all English instruction. This alternative theory suffers the same problem as does the facilitation hypothesis: there is no consistent evidence for it (or against it) in the bilingual education literature. Every study that supports the facilitation hypothesis contradicts the time on task hypothesis *and vice versa*.

Porter's argument does not stand up under scrutiny. Porter (1990a) says "Education research, we have seen, makes the compelling case for time on task as the most important single

determinant of success in learning *anything*” [emphasis in the original] (p. 125). Later I show that this overly simplified time on task model is wrong; it does not hold at certain stages of learning unless some conditions, of which Porter seems to be blissfully ignorant, are specified. Porter mentions the time on task hypothesis twice before that citation. The first is “My own observations of the most effective teaching of [LEPs]...have led me to the following conclusions... Providing as much time as possible for [LEP] students to use and practice English... produces the greatest success in learning English” (p. 83). The second follows a discussion of the Canadian immersion studies: “the more time spent learning a language, the better you do in it.” These two comments hardly justify her sweeping claim that research has solidly established her pet theory as the most important principle in learning anything. Moreover, Karweit (1982), in a review of the time on task literature concludes the time on task effect, while real, is small and its importance has been greatly overblown.

Porter replaced bilingual education in Newton, Mass. with a poorly described all English program. She repeatedly claims that her program is far superior to bilingual education programs. However, there is no evidence in the book to support her sweeping claims. In place of test scores, dropout rates, or other empirical indicators, Porter offers her personal opinion that she is right and the bilingual educators are wrong. Indeed, the book so lacks empirical evidence about the effects of the Newton program that there is no reason to think it teaches English any better than would have happened if there were no special program at all.

While Porter offers no empirical evidence that the Newton program succeeded, she does present a highly selective discussion of some of the research on teaching language-minority children. Porter’s discussion of the Fairfax County’s ESL program illustrates the biased way she approaches the literature. Referencing one report from the Fairfax school system, Porter concludes “These evaluations (sic) portray a program that is working well for its students” (p. 148). Porter makes no mention of Collier’s extensive evaluation of the Fairfax program which recommended replacing it with a bilingual education program (Collier 1987a,b).

She summarizes a report from El Paso as finding that an all English immersion program was superior to bilingual education programs. The El Paso report has no such finding. What Porter

describes as an all English immersion program in El Paso is in fact a Spanish-English dual immersion program.

Porter describes an all English “structured immersion program” in San Diego. I invented the term structured immersion (Baker & de Kanter 1981), and I can say with confidence there is not now, nor was there ever, a structured immersion program in San Diego. San Diego uses a dual immersion bilingual education program model. The “San Diego Two-Way bilingual Program... is initially presented 90% in Spanish and 10% in English, gradually arriving at a balance of 50% Spanish and 50% English” (Herbert 1987).

Like El Paso, San Diego has a dual-language bilingual program. There is evidence that the extensive bilingual education program worked better than the typical bilingual education program. Like El Paso, the results of the San Diego study are the opposite of what Porter reported. The results of the San Diego study argue, if anything, for more bilingual education programs, not fewer as Porter maintains.

Porter describes a program in Paramount, CA. in which she says students in an English immersion program “show a far higher rate of English-language learning than the children in the Spanish bilingual program” (p. 152). Not recognizing this study, I consulted Porter’s reference for it, a report prepared for the Reagan Administration (National Advisory and Coordinating Council on Bilingual Education 1987) by a group of educators (including Porter herself) opposed to bilingual education programs. There was no data in that report either, nor was there any reference to any study supporting Porter’s claim.

I have test scores from this program for the year in which Porter wrote her book. The school system did have both a bilingual education program and an all English immersion program. Of 12 comparisons between the two programs (reading, language, math, and total scores for three grades), the bilingual education program students had the higher score for 8 of the 12 comparisons. The evidence points to a conclusion just the opposite of Porter’s. In short, Porter’s time-on-task theory of L2 learning is on as shaky grounds as is the facilitation hypothesis.

Because these theories were so poorly supported by the empirical evidence outside the Ramírez study, Ramírez et al. should not have relied so heavily on them to guide their analysis. In short, the analysis should have been tempered with crass empiricism, which shows that Ramírez et al. found an immediate superiority for

bilingual education programs, something contrary to the predictions of both the facilitation hypothesis and the time-on-task hypothesis.

For the Ramírez et al. data, the facilitation hypothesis implies superiority for structured immersion in the early years. That did not happen. The facilitation hypothesis implies a shift in advantage from immersion in the early years to bilingual education later on. That did not happen. The time on task hypothesis implies both an initial superiority for immersion and an increase in that superiority over time. Neither happened. The facts are that bilingual education programs were superior to all English instruction in the early stages of learning English. Since neither theory fits the facts, both theories must be invalid. Neither theory correctly explains how English is learned as a second language.

This is a strong conclusion, and would be more tentative were it not supported by considerable additional empirical evidence discussed earlier.

The time on task theory of second language learning is also compellingly contradicted by the evidence. One outstanding feature of the Ramírez study is that it is one of the few studies to actually measure the use of English in the classroom. Classroom observations were made for two periods each year. Table 7 shows teacher use of English in the three programs. There are dramatic differences in the use of English in the classroom, but it seems not to affect learning English. Clearly, both the facilitation and the time on task hypotheses are contradicted by these data.

Table 7
Percent of teacher speech in English

Grade	All English	Bilingual A	Bilingual B
K	98.5	65.8	9.3
1	97.3	69.1	32.9
2	98.2	74.5	30.3
3	99.0	80.3	50.6
4	99.8	97.3	55.3
5	NA	NA	63.6
6	NA	NA	80.3

Source: Ramírez et al. (1991).

Empirical findings are considerably strengthened if they are placed in a theoretical framework. I have argued that the facts in Ramírez et al. contradict the theoretical framework of the study, indeed, that the theory misled the authors as to what findings were

or were not important. The question then arises, can the initial superiority of bilingual education over all-English instruction be placed in a theoretical context? The next section shows how.

An Alternative hypothesis

There is a well established phenomenon in experimental psychology that accounts for all the results discussed. It manifests itself in the classroom as attention span. Students, especially young students, can only focus their attention on any one lesson for relatively short periods of time. Teachers know they have to change the lesson every 10-30 minutes or learning stops. A child who does not speak English who is put in an all English environment for 6-7 uninterrupted hours of English will quickly stop learning English; the "lesson", i.e., 6 hours of English, exceeds his attention span.

This phenomenon has long been known in psychology as the effects of massed versus spaced trials or continuous versus spaced practice. It is well established that massed and spaced learning trials have strong effects on learning. One classic demonstration of the spacing effect (Lyon 1917) looked at the average length of time it took to memorize nonsense syllables in lists of different length, comparing continuous practice to once a day practice. For a list of 40 nonsense syllables, the average learning time per syllable was four times longer for continuous practice than when the list was practiced just once a day. This is a particularly interesting study because of the similarity between nonsense syllable learning and acquiring vocabulary in L2.

Duncan (1951) studied the acquisition of keeping a pen point on a moving target. One group of subjects practiced for the entire time allocated for learning. The other group was periodically interrupted for rest periods so that they were resting for two thirds of the whole practice session. The group with less practice (more rest) learned better. Kientzle (1946) found that printing upside down was learned better with a rest of seven days between trials than with continuous practice or with as much as a 20 second rest between trials. As indicated by the dates of these studies, the superiority of spaced learning has been known for a long time. However, educational research, theory, and practice are notable in their failure to recognize and apply this phenomenon (Demster 1988).

The probable explanation for the superiority of spaced practice over continuous practice is that it takes time for memory process to work. A constant barrage of new material to learn overloads

memory and interferes with learning. Rest, or doing something else, between learning sessions gives memory the time it needs to operate and more efficient learning results. Consider L1 learning in the school-aged child. This child already knows most of the words the teacher uses on any particular day. The few new words to be learned are interspersed with long periods of no learning of new words -- rest. *Language development ordinarily takes place in a setting of spaced practice, not continuous practice, and the learner makes rapid progress.*

However, the situation is different for the young child beginning to learn an L2. At first, since all words are new, exposure to the L2 creates a situation of continuous learning rather than spaced learning. He will make more progress learning the L2 if rest (a break in the exposure to the new language) is introduced into the constant stream of exposure to the unknown language.

How can breaks be taken in exposure to a new language? By changing to teaching in the language the child already knows. Eventually, enough of the L2 will be learned so that the second language learner is in the same situation as the native speaker: Rest between new words is built into the lesson because the teacher mainly uses words already learned.

This theory predicts a pattern where bilingual education programs, because they provide the much needed rest from constant exposure to the new language, produce better learning at the early stages of developing L2, but later on instruction entirely in the L2 works as well or better than bilingual education. This is precisely the pattern Burkheimer, et al. (1990) found in their extensive study.

In short, it is a waste of time to try to teach anything, including English, for 6 solid hours to little kids. The longer a lesson goes on, the less efficient learning becomes. Out of the 6-7 hours of the elementary school day, considerably less than that total amount of time is all that is needed for effectively learning English by children who do not speak English. By using the native language to introduce periods of rest, learning English is made more effective, hence the superiority of bilingual education programs over all English programs in the early stages of learning English.

The Ramírez et al. analysis is consistent with the spacing effect theory of the effects of bilingual education programs, but it does not support the facilitation effects.

In a sense, the time on task and spaced learning effects work against each other in the early stages of L2 learning. Later, they

reinforce each other as the rest periods built into further learning of a language whose basics are already learned generate the spacing effect allowing longer lessons that then reveal the time on task effect.

Conclusion

Ramírez et al. seem to have overlooked the importance of their strongest finding, the immediate benefits of bilingual education programs, while overemphasizing the importance of a speculative and ultimately invalid analysis of late-exit programs because they placed too much reliance on poorly formed theories of how LEP children learn English.

Ramírez et al. gave short shrift to their most important finding—that transitional bilingual education programs are superior to all-English instruction in the early stages of learning English, most likely because the facilitation hypothesis predicted long run, not short run, effects and the alternative hypothesis of the opponents of bilingual education programs predicted continuous advantages of all-English instruction.

Ramírez et al.'s arguments favoring late-exit programs are grossly speculative and contradicted by their own data. They are interesting only because they provide a degree of fit with the facilitation hypothesis. If the facilitation hypothesis were a decent theory, we might be able to overlook the weaknesses of Ramírez et al.'s analysis, but the facilitation hypothesis is so lacking empirical support in the literature that it merits no further consideration.

Ramírez et al. provide strong disproof of both the time-on-task, "teach 'em in English" theory and the facilitation hypothesis. That bilingual education programs have immediate advantages over all-English instruction is a finding not unique to Ramírez et al. It is found in over a dozen methodologically sound studies, and there is a theoretical explanation as to why that is the way bilingual education programs work.

References

- American Institute for Research. 1975. *Alice Independent School District in Identification and description of exemplary bilingual education programs*. Palo Alto, CA. Author.

- Baker, K. 1984. Ideological bias in bilingual education research. Paper presented at the Annual meeting of the American Educational Research Association, San Francisco.
- Baker, K. 1990. Bilingual education's 20-year failure to provide civil rights protection for language-minority students. *Children at risk: Poverty, minority status, and other issues in educational equity*. (Eds. A. Barona & E. Garcia). Washington, DC: National Association of School Psychologists.
- Baker, K. & A. de Kanter. 1981. *The effectiveness of bilingual education programs: A review of the literature*. Final draft report. Washington, DC: U.S. Department of Education.
- Baker, K. & A. de Kanter. 1983. An answer from research on bilingual education. *American Education*. 56(4), 157-169.
- Baker, K. & C. Rossell. 1987. An implementation problem: Specifying the target group for bilingual education. *Education Policy Analysis*. 1(2), pp. 249-270, p. 263.
- Barkley, L. 1983. Using Spanish as the language of instruction with M-A headstart children. *Journal of Perceptual & Motor Skills*, 56, 359-366.
- Burkheimer, Jr., G., Conger, A., Dunteman, G., Elliott, B. & K. Mowbray. 1990. *Effectiveness of services for language-minority limited English proficient students*. Research Triangle Institute. Raleigh-Durham, NC.
- Cardoza, D., Sánchez, A. & R. Mendoza. 1985. Attitudes toward bilingual education and foreign language instruction among four ethno-linguistic groups. Los Alamitos, CA: NCBE, April.
- Collier, V. 1987a. *The effect of age on acquisition of a second language for school*. National Clearinghouse for Bilingual Education.
- Collier, V. 1987b. Age and rate of acquisition of second language for academic purposes. *TESOL Quarterly*. 214, 617-641.

- Cummins, J. 1978. Educational implications of mother tongue maintenance in minority language groups. *Canadian Modern Language Review*. (34), 395-416.
- Cummins, J. 1981. The role of primary language development in promoting educational success for language-minority students. In California State Dept. of Education (Comp.), *Schooling and language-minority students: A theoretical framework*. Los Angeles: UCLA Evaluation, Dissemination, & Assessment Center.
- Cummins, J. 1985. The construct of lang proficiency in bilingual education. *Perspectives on Bilingualism and Bilingual Education*. pp. 209-231. Georgetown University Press.
- Demster, F. 1988. The spacing effect: A Case study in the failure to apply the results of psychological research. *American Psychologist*. 43(8), 627-634.
- Dulay, H. & M. Burt. 1978. From research to method in bilingual education. Georgetown University roundtable on language and linguistics. Washington, DC: Georgetown University.
- Duncan, C. 1951. The effect of unequal amounts of practice on motor learning before and after rest. *Journal of Experimental Psychology*. 42, 257-264.
- Engle, P. 1975. The use of the vernacular language in education. Bilingual Education Series 2. Washington, DC: The Center for Applied Linguistics.
- Herbert, C. 1987. *San Diego Title VII Two-Way Bilingual Education Program*. San Diego: San Diego Public Schools. (May 22).
- Holland, R. 1986. *Bilingual education: Recent evaluations of local school district programs and related research on second-language learning*. Report No. 86-611. Washington, DC: Congressional Research Service.
- Karweit, N. 1982. *Time on Task: a Research Review*. Washington, DC: National Commission on Educational Excellence.

- Kientzle, J. 1946. Properties of learning curves under varied distribution of practice. *Journal of Experimental Psychology*. 36, 187-211.
- Legareta, D. 1979. The effects of program models on language acquisition by Spanish speaking children. *TESOL Quarterly*. 13, 521-534.
- Lyon, D. 1917. *Memory and the learning process*. Baltimore: Warwick and York.
- McConnell, B. 1980. Effectiveness of individualized bilingual instruction for migrant students. Doctoral Dissertation, Washington State University.
- Morgan, J. 1971. The effects of bilingual instruction on them English language arts achievement of first grade children. Doctoral Dissertation, Northwestern State University of Louisiana.
- National Advisory and Coordinating Council on Bilingual Education (NACCBE), U. S. Department of Education. 1987. *Promising Practices in Bilingual Education: School Visit Project*. Eleventh Annual Report to Congress. Washington, D.C.: Government Printing Office, 29-71.
- Pelavin, S., Celebuski, C., Fink, L. & K. Baker. 1988. Determining which non-English background students need special language services. Paper presented at the annual meeting of the American Education Research Association, San Francisco. (April).
- Peterson, M., Berry, D., Abbott, S., Kruvant, C., Sundusky, N., Chow, C. & G. Ortega. 1976. *Assessment of the Status of Bilingual Vocational Training: Review of the Literature*. ERIC Doc. No. ED 131 683.
- Porter, R. 1990a. *Forked Tongue*. New York: Basic Books.
- Porter, R. 1990b. Reflections on the politics of bilingual education. *Journal of Law and Politics*, VI(3), 573-599.

- Ramírez J., S. Yuen, D. Ramey & D. Pasta. 1991. *Final Report: Longitudinal study of structured English immersion strategy, early-exit and late-exit bilingual education programs for language-minority children*. (Vol. I) (Prepared for U.S. Department of Education). San Mateo, CA: Aguirre International. No. 300-87-0156.
- Ramírez, J., Pasta, D., Yuen, S., Ramey, D. & D. Billings. 1991. *Final report: Longitudinal study of structured English immersion strategy, early-exit and late-exit bilingual education programs for language-minority children*. (Vol. II) (Prepared for U.S. Department of Education). San Mateo, CA: Aguirre International. No. 300-87-0156.
- Ravitch, D. 1983. *The Troubled Crusade*. New York: Basic Books.
- Rossell, C. & J. Ross. 1986. The social science evidence on bilingual education. *Journal of Law and Education*, 15, 385-419.
- Rossell, C. & K. Baker. 1988. Selecting and exiting students in bilingual education programs. *Journal of Law and Education*. 17(4), 589-23.
- Rotberg, I. 1982. Federal policy in bilingual education. *American Education*. 52(2), 30-0.
- Skutnabb-Kangas, T. 1979. *Language in the process of cultural assimilation*. Washington, DC: National Clearinghouse for Bilingual Education.
- Skutnabb-Kangas, T. & P. Toukomaa. 1976. *Teaching migrant children's mother tongue and learning the language of a host country*. Helsinki: The Finnish National UNESCO Commission.
- Troiike, R. 1983. Bilingual Si! *Principal*. (January Issue, 8).
- Troiike, R. 1978. Research evidence for the effectiveness of bilingual education. *NABE Journal*, 3(1), 13-4.

Willig, A. 1985. A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, 55, 269-17.

Yates, J. & A. Ortiz. 1983. Baker & de Kanter review: Inappropriate conclusions on the efficacy of bilingual education. *Journal of the National Association for Bilingual Education*. 7(3), 75-84.

Zirkel, P. 1972. *An evaluation of the effectiveness of selected experimental bilingual education programs in Connecticut*. Hartford, CN: Dept. of Education.