

A Meta-Analysis of the Effectiveness of Bilingual Education

by  
Jay P. Greene  
Assistant Professor of Government  
University of Texas at Austin  
March 2, 1998

Sponsored by  
The Tomas Rivera Policy Institute  
The Public Policy Clinic of the Department of Government, University of Texas at Austin  
The Program on Education Policy and Governance at Harvard University

## **Introduction**

The voters in California are being asked to consider an initiative this June that would ban the use of foreign languages in the instruction of younger children with limited English proficiency. Both advocates and opponents of this initiative claim that scholarly research supports their case, but their reading of the literature is often selective, exaggerated, and distorted. With the sponsorship of the Tomas Rivera Policy Institute, the Public Policy Clinic of the University of Texas' Government Department, and Harvard University's Program on Education Policy and Governance, I have conducted a systematic, statistical review of the literature on the effectiveness of bilingual education.<sup>1</sup> With this technique known as meta-analysis to summarize the scholarly research, I find that children with limited English proficiency who are taught using at least some of their native language perform significantly better on standardized tests than similar children who are taught only in English. In other words, an unbiased reading of the scholarly research suggests that bilingual education helps children who are learning English.

## **Estimated Benefit of Bilingual Education**

This conclusion is based on the statistical combination of eleven studies that meet minimal standards for the quality of their research design from a total of seventy-five studies reviewed. These eleven studies include standardized test score results from 2,719 students, 1,562 of whom were enrolled in bilingual programs, in thirteen different states. The estimated benefit of using at least some native language in instruction on all scores measured in English is .18 of a standard deviation on standardized tests. The average student in these bilingual programs was tested in third grade after two years of bilingual instruction. Bilingual programs produce .21 of a standard deviation improvement on reading tests and .12 of a standard deviation improvement on math tests measured in English. The gain in all test scores measured in Spanish is .74 of a

---

<sup>1</sup> Research assistance was provided by Luis Guevera. I also want to thank Larry Bernstein, Elsa DelValle-Gaster, Rudy de la Garza, Charles Glenn, Aleza Greene, Kenji Hakuta, Stephen Krashen, Michael Kwiatkowski, Tse Min Lin, Gary Orfield, Harry Pachon, Paul Peterson, Joel Spalter, and Ann Willig for

standard deviation. All of these gains, except for math, are statistically significant, meaning that they are unlikely to have been produced by chance. (See Table 1 for summary of results.)

### **Interpreting Standard Deviations**

To put the size of this benefit in perspective, the gap between the scores of minority and white students on standardized tests nationwide is about 1 standard deviation. The estimated benefits of bilingual education are also comparable to the improvements produced by the school choice program in Milwaukee that I have studied, where students gained between 1/3 and 1/2 of a standard deviation after four years of participation. (Greene et al, 1997) Education researchers generally consider a gain of .1 standard deviation as slight, .2 or .3 of a standard deviation as moderate, and .5 of a standard deviation as large. (Hanushek, 1996; Hedges and Greenwald, 1996)

In more concrete terms, we can imagine two identical students with limited English proficiency who enter first grade scoring at the 30<sup>th</sup> percentile on the reading component of the Iowa Test of Basic Skills (ITBS), meaning that 70% of those who take the same test in first grade perform better than they do. After two years in which one student was in a bilingual program and the other student was in an English-only program, the bilingual student would be performing about 1/5 of a standard deviation better than the English-only student on the ITBS reading test. If the English-only student scored at the 26<sup>th</sup> percentile at the end of those two years, we would expect the bilingual student to score at the 34<sup>th</sup> percentile. (See Figure 1) The English-only student would be five months behind grade level, while the student in the bilingual program would be only two months behind grade level. According to this hypothetical, students in bilingual programs receive the equivalent of roughly three additional months of learning over a two-year period compared to similar students in English-only programs.

---

their helpful comments. In addition to helpful suggestions, Christine Rossell and James Yates provided some of the harder to find studies.

### **Estimated Benefits of Bilingual Education from Random Assignment Studies**

Random assignment to treatment and control groups, as in medical experiments, is the highest quality research design because it increases the confidence in the conclusion that any differences between the groups after a period of treatment can be attributed to that treatment. The results from the five studies in which subjects were randomly assigned to bilingual and control programs favor bilingual education even more strongly. The estimated benefit of bilingual programs on all test scores in English according to these studies with random assignment is .26 of a standard deviation. The positive effect on reading scores is .41 of a standard deviation among the studies with random assignment. And the improvement in scores measured in Spanish is .92 of a standard deviation in the studies with random assignment to treatment and control groups. All of these estimated benefits of bilingual education from studies with random assignment are extremely unlikely to have been produced by chance (the odds are fewer than 1 in 100). (See Table 2 for summary of results) The fact that the studies of bilingual programs with random assignment, the highest quality research design, have even stronger results greatly increases the confidence in the conclusion that bilingual education positively affects educational attainment.

A meta-analysis of the 11 studies that meet minimal standards for the quality of their research design as well as the 5 highest quality studies based on random assignment show positive, statistically significant, benefits for bilingual education. The results of this meta-analysis are similar to the meta-analysis conducted by Ann Willig in 1985 based on the Baker and de Kanter review of the literature in 1981. While few acceptable-quality studies have been conducted in the intervening years, the conclusions that Willig drew from the literature are still true today: the evidence that is available suggests that native language instruction has a significant, positive impact on children learning English.

### **Method for Selecting Studies and Computing Results**

The eleven studies included in this meta-analysis are drawn from a list of 75 “methodologically acceptable” studies compiled by Christine Rossell and Keith Baker, two vocal

critics of bilingual education, in a 1996 literature review (Rossell and Baker 1996). The Rossell and Baker list is used as the pool of studies examined for this meta-analysis for a few reasons despite the potential for bias in their selections. First, Rossell and Baker claim to have selected their methodologically acceptable studies based on criteria that I believe are reasonable. To be acceptable the studies had to:

- 1) compare students in a bilingual program to a control group of similar students
- 2) differences between the treatment and control groups had to be controlled statistically or assignment to treatment and control groups had to be random
- 3) results had to be based on standardized test scores in English, and
- 4) differences between the scores of treatment and control groups had to be determined by applying appropriate statistical tests.

In addition to these requirements this meta-analysis only included studies that measured the effects of bilingual programs after at least one academic year. Bilingual programs were defined as ones in which students with limited English proficiency are taught using at least some of their native language. An appropriate control group was one in which students were taught only in English. If students were not assigned to treatment and control groups randomly, adequate statistical controls for this non-random assignment was defined as requiring controls for individual previous test scores as well as at least some of the individual demographic factors that influence test scores (e.g. family income, parental education, etc...). Rossell and Baker identify 72 studies that they say meet these standards (although there are actually 75 citations listed under their heading for acceptable studies in a mimeo provided by Rossell).

Second, critics of Rossell and Baker's literature review have not offered additional studies that meet the above criteria. Stephen Krashen, a vocal proponent of bilingual education for example, has instead suggested that the standards are too strict and has proposed that Rossell and Baker include additional studies favorable to bilingual education even though they do not meet the criteria (Krashen 1996). Here I have to agree with Rossell and Baker that their standards

are reasonable and reject considering Krashen's additional studies. The inability of others to advance the names of more studies that meet Rossell and Baker's criteria lends credence to the assumption that their list is a comprehensive pool from which to select acceptable studies for a meta-analysis.

Unfortunately, only 11 of the 75 studies identified as acceptable by Rossell and Baker actually meet their own criteria for an acceptable study. Fifteen of the studies duplicate the evaluations found in some of the remaining 60 studies. That is, 15 of the 75 are separately released reports of the same programs by the same authors that are already included in Rossell and Baker's list. Where appropriate I combine results so that each remaining observation represents an independent evaluation of a program. Despite our best efforts, an additional 5 studies in Rossell and Baker's list could not be found. While Christine Rossell was very helpful in locating some of the more difficult to find studies, she did not have these 5 nor were they available from the library at the University of Texas, which has one of the world's largest collections. (See the annotated bibliography for a list of studies and the reasons for their exclusion or inclusion in the meta-analysis).

Of the remaining 55 studies, 3 are excluded because they are not evaluations of bilingual programs. One is about "direct instruction"(Becker 1982) and makes no mention of foreign language learning. Another is a list of exemplary bilingual programs (Campeau 1975), not an evaluation of programs. And yet another is primarily about the effects of retention (being held back a grade) (Webb 1987).

An additional 14 studies are excluded because they do not have adequate control groups. In most of these studies both the treatment and control groups receive bilingual instruction, meaning that all students are taught in both their native language and in the target language in varying amounts. I only include in the meta-analysis studies that compare bilingual instruction (meaning the use of at least some native language in instruction) to "English-only" instruction. There are several reasons for this choice. First, comparing the use of some native language to

English-only instruction is the clearest division possible in the literature. Program labels, such as transitional bilingual education, English as a second language, immersion, submersion, and maintenance bilingual education, have no consistent meaning in the evaluations, nor are the detailed features of many programs fully described. The only division of programs that can accurately and consistently be applied is whether native languages are used in instruction or not.

Second, the most policy-relevant question and the issue raised by the initiative in California is whether it is desirable to ban the use of native language instruction in the education of younger students with limited English proficiency. The question is not whether it is better to use a modest amount of native language versus a large amount, nor is the issue whether it is better to have children in bilingual programs for a short versus long time. Thus only studies that speak to the policy-relevant issue of comparing bilingual to English-only instruction are included in this meta-analysis. In addition, it is not possible to extrapolate results from studies that compare different amounts or lengths of bilingual instruction to whether bilingual instruction is desirable at all. Similarly, if one wanted to know whether acetaminophen was effective in treating headaches, it would be incorrect to infer an answer from a study that gave different doses to a treatment and control group. Giving 500 mg of acetaminophen to one group may cure their headaches and giving 10,000 mg to another group may kill them. It would then be wrong to extrapolate from these results to the claim that acetaminophen is harmful in any dose. The only way to evaluate whether the use of any native language instruction is harmful or helpful is to compare students who receive any bilingual instruction to those who are taught only in English.

Of the remaining 38 studies, 2 are excluded because they measure the effects of bilingual programs after an unreasonably short period of time. One study evaluates a program after 7 weeks of bilingual instruction for 35 minutes a day (Barclay 1969). The other evaluates a program after 10 weeks (Layden 1972). Every study included in the meta-analysis measures effects after at least one academic year (about 40 weeks). While the requirement that studies evaluate the effects of bilingual programs after at least one academic year was not one of Rossell

and Baker's original criteria for identifying acceptable studies, this is a reasonable standard to add. To make the analogy to headache cures again, measuring bilingual programs after 7 or 10 weeks is like measuring the effects of aspirin after 1 minute. No valuable information can be gained from evaluating such a short period of treatment.

An additional 25 studies are excluded because they inadequately control for the differences between students assigned to bilingual programs and students assigned to English-only control groups. If students are randomly assigned to these two groups, then no controls are necessary and one can place high confidence in the results. But when students are not randomly assigned, it is necessary to control statistically for the differences between the groups that may affect their future performance. Three of these 25 studies make no effort to control for the differences between bilingual and English-only students (Curiel 1979, Valladolid 1991, and Yap 1988). Some studies do not control for individual characteristics of students but instead match aggregate characteristics of students in a program with aggregate characteristics of a control group (see Stebbins 1977 for example). Without controlling for individual level factors these studies suffer from the "ecological fallacy" where the uncontrolled individual factors that contribute to improved performance seriously bias the aggregate results. Most of the other 25 studies excluded for inadequate background controls, however, only control for test scores taken earlier or IQ test scores when estimating the effects of bilingual instruction.

For these to be adequate controls for the differences between the groups, one would have to assume that the rate of test score gains, absent any treatment, would be the same for students with different initial test scores or different IQ's. Yet considerable evaluation research (See Campbell and Erlebacher 1970) has shown that students who begin with different test scores often have different rates of growth in their test scores. For example, a student with low initial scores may have those scores in part because she is poor and does not have parents involved in her education. Those same factors that contributed to the low initial score may continue to reduce her educational progress in the future. Unless one controls for the differences in initial scores, as

well as some of the important factors that produce those different scores, evaluations of educational progress are likely to be significantly biased.

The remaining 11 studies that are included in the meta-analysis consist of 5 studies in which students are randomly assigned and 6 in which there is non-random assignment but some effort to control for the individual background characteristics as well as test scores that separate those in bilingual and English-only programs. A single, average effect size was calculated for each study for each subject area and for all tests in English and Spanish. The effect sizes were standardized and adjusted for their sample size into corrected units of standard deviations known as Hedge's  $g$ . The mean of the 11 Hedge's  $g$ 's was then computed as the reported estimated effect of bilingual programs. A single, average  $z$ -score (a statistical measure of confidence in the estimated effect) was also calculated for each study for each subject area and for all tests in English and Spanish. The  $z$ -scores were combined by adding them and then dividing by the square root of the number of studies to compute a combined  $z$ -score.  $P$  values can then be calculated from the combined  $z$ -score. These techniques are described at greater length in Rosenthal 1991 and Cook, et al 1992.

### **Differences between These Results and Rossell and Baker's Results**

It is important to note that the positive estimated effects of bilingual education in this meta-analysis are not simply a product of the selection of these 11 acceptable studies. Of the 38 studies that evaluate bilingual versus English-only programs in Rossell and Baker's list, 21 have an average positive estimated effect and 17 have an average negative estimated effect. Simply counting positive and negative findings, however, is less precise than a meta-analysis because it does not consider the magnitude or confidence level of effects. In addition, once we include unacceptable studies from Rossell and Baker's list we would also have to consider the methodologically unacceptable studies advanced by Krashen and other supporters of bilingual education. Nevertheless, even when studies with inadequate background controls and short

measurement periods only from Rossell and Baker's list are included, we still find that the scholarly literature favors the use of native language in instruction.

Rossell and Baker report a different number of positive and negative studies for a few of reasons. First, they include in their report studies that are redundant with other studies, not available, not evaluations of bilingual programs, and do not have English-only control groups. Second, they do not apply any consistent rule for classifying studies as positive or negative. For example, Ramirez 1991 is classified as showing "no difference" despite having significant, positive effects for bilingual instruction in reading. Similarly, Education Operation Concepts 1991 is classified as showing that bilingual education has a negative effect on reading scores despite having no statistically significant effects (and the average effect is actually positive, not negative). One of the advantages of meta-analysis is that it forces one to be consistent in summarizing other research. Third, there are some studies in their categories of positive and negative studies that are not found in their list of acceptable studies (such as Olesini 1971 and Elizondo 1972). It is clear that Rossell and Baker's review of studies is useful as a pool for a meta-analysis, but the lack of rigor and consistency in how they classify studies and summarize results prevent their conclusions from being reliable.

### **Conclusion**

While it would be desirable to have a meta-analysis based on a greater number of studies, the unfortunate reality is that the vast majority of evaluations of bilingual programs are so methodologically flawed in their design that their results offer more noise than signal. Adding seriously flawed studies would bias the results of this meta-analysis in ways that are nearly impossible to predict or correct. In addition, including studies that do not meet minimal criteria would require identifying the entire universe of inadequate studies and including all or a random sample of those studies in a meta-analysis. The incredible amount of effort that would require is not justified given the low amount of information that could be gained. Focusing on studies that meet certain "bright-line" criteria, such as all studies that control for individual background

characteristics as well as pretest scores or on the smaller group of studies based on random assignment, provides an unbiased sample of studies that can offer useful information on the effects of bilingual education. Despite the relatively small number of studies, the strength and consistency of these results, especially from the highest quality randomized experiments, increases confidence in the conclusion that bilingual programs are effective at increasing standardized test scores measured in English.

The limited number of useful studies, however, makes it difficult to address other important issues, such as the ideal length of time students should be in bilingual programs, the ideal amount of native language that should be used in instruction, and the age groups in which these techniques are most appropriate. It is possible that the individual needs of students are so varied that there may be no simple set of ideal policies. But if we want to learn more about how to develop public policy that is most effective at addressing the needs of students with limited English proficiency, we need to conduct a series of experiments in which students are randomly assigned to different types of programs. These randomized experiments yield the clearest and most precise information to help guide policymaking. The results from the 5 randomized experiments examined here clearly suggest that native language instruction is useful. We need additional randomized experiments to determine how best to design those bilingual programs.

Table 1: Results from the Meta-Analysis of the Effects of Bilingual Education

	All tests in English	Reading (in English)	Math (in English)	All tests in Spanish
Benefit of Bilingual Programs in Standard Deviations (Hedge's g)	.18	.21	.12	.74
z-score	2.41	2.46	1.65	3.53
p-value <	.05	.05	.10	.01

Table 2: Results from the Meta-Analysis of the Effects of Bilingual Education for Studies with Random Assignment to Bilingual and Control Programs

	All tests in English	Reading (in English)	Math (in English)	All tests in Spanish
Benefit of Bilingual Programs in Standard Deviations (Hedge's g)	.26	.41	.15	.92
z-score	2.71	3.47	1.25	5.21
p-value <	.01	.01	.21	.01

Table 3: Summary of Results from Studies Included in Meta-Analysis

Study	English		Reading		Spanish		Treatment	Control	Random Assignment
	ES	Z	ES	Z	ES	Z	N	N	Yes/No
Bacon, 1982	.79	2.39	.68	2.07	NA	NA	18	18	No
Covey, 1973	.34	2.94	.74	4.87	NA	NA	86	89	Yes
Danoff, 1977	-.03	-.39	-.12	-1.50	NA	NA	955	523	No
Huzar 1973	.18	.83	.18	.83	NA	NA	43	43	Yes
Kaufman, 1968	.20	.72	.20	.72	1.65	6.05	43	31	Yes
Plante, 1976	.52	1.34	.52	1.34	1.09	2.89	16	12	Yes
Powers, 1978	.001	.01	-.33	-1.53	NA	NA	44	43	No
Ramirez, 1991	.01	.08	.12	.73	NA	NA	88	160	No
Rossell, 1990	-.01	.03	-.05	-.20	NA	NA	174	173	No
Rothfarb, 1987	.05	.24	NA	NA	.01	.09	70	49	Yes
Skoczylas, 1972	-.05	-.18	.13	.46	.20	.68	25	25	No

ES = Average effect size measured in standard deviations (Hedge's g)

N = Largest number of subjects in any analysis in the study. For Huzar, 1973 and Rossell, 1990 the number of subjects in the treatment and control groups had to be estimated by halving the total reported sample.

## Annotated Bibliography

### Methodologically Acceptable Studies Included in the Meta-Analysis

Bacon, Herbert L., Kidd, and Gerald D., et al. 1982. "The Effectiveness of Bilingual Instruction with Cherokee Indian Students." Journal of American Indian Education. February. pp. 34-43.

Covey, D. D. 1973. "An Analytical Study of Secondary Freshmen Bilingual Education and its Effects on Academic Achievement and Attitudes of Mexican American Students." Ph.D. dissertation. Arizona State University.  
Random assignment.

Danoff, Malcom N., Arias, B.M., Coles, Gary J., and Others. 1977a. Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program. American Institutes for Research. Palo Alto.

Huzar, Helen. 1973. "The Effects of an English-Spanish Primary Grade Reading Program on Second and Third Grade Students." M.Ed. thesis. Rutgers University.  
Random assignment.

Kaufman, Maurice. 1968. "Will Instruction in Reading Spanish Affect Ability in Reading English?" Journal of Reading. Volume. 11. pp. 521-527.  
Random assignment.

Plante, Alexander J. 1976. A Study of Effectiveness of the Connecticut "Paring" Model of Bilingual/Bicultural Education. Connecticut Staff Development Cooperative. Hamden.  
Random assignment.

Powers, Stephen. 1978. "The Influence of Bilingual Instruction on Academic Achievement and Self-Esteem of Selected Mexican American Junior High School Students." Ph.D. dissertation. University of Arizona.

Ramirez, J. David, Pasta, David J, Yuen, Sandra, Billings, David K., and Ramey, Dena R. 1991. Final Report: Longitudinal Study of Structural Immersion Strategy, Early-Exist, and Late-Exist Transitional Bilingual Education Programs for Language-Minority Children. Aguirre International (report to the U.S. Department of Education): San Mateo.

Rossell, Christine H. 1990. "The Effectiveness of Educational Alternatives for Limited-English-Proficient Children.: in Imhoff, Gary. (ed.). Learning in Two Languages. Transaction Publishers. New Brunswick.

Rothfarb, Sylvia H., Ariza, Maria J. and Urrutia, Rafael. 1987. Evaluation of the Bilingual Curriculum Content (BCC) Project: A Three-Year Study, Final Report. Office of Educational Accountability. Dade County.

Skoczylas, Rudolph V. 1972. "An Evaluation of Some Cognitive and Affective Aspects of a Spanish Bilingual Education Program". Ph.D. dissertation. University of New Mexico.

### Studies Excluded Because They are Redundant

Ariza, Maria. 1988. "Evaluating Limited English Proficient Students' Achievement: Does Curriculum Content in the Home Language Make a Difference?" Paper presented at the April meetings of the American Educational Research Association. New Orleans.  
Redundant with Rothfarb et al, 1987.

Barik, Henri, and Swain, Merrill. 1978. Evaluation of a Bilingual Education Program in Canada: The Elgin Study Through Grade Six. Commission Interuniversitaire Suisse de Linguistique Appliquee. Switzerland.  
Redundant with Barik et al 1977.

Cohen, Andrew D., Fathman, Ann K., and Merino, Barbara. 1976. The Redwood City Bilingual Education Report, 1971-1974: Spanish and English Proficiency, Mathematics, and Language-Use Over Time. Ontario Institute for Studies in Education. Toronto.  
Redundant with Cohen 1975.

Curiel, Herman, Stenning, Walter, and Cooper-Stenning, Peggy. 1980. "Achieved Ready Level, Self-Esteem, and Grades as Related to Length of Exposure to Bilingual Education." Hispanic Journal of Behavioral Sciences. Vol. 2. pp. 389-400.  
Redundant with Curiel, 1979.

Danoff, Malcom N., Coles, Gary J., McLaughlin, Donald H., and Reynolds, Dorothy J. 1977b. Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs, Volume I: Study Design and Interim Findings. American Institutes for Research. Palo Alto.  
Redundant with Danoff et al 1977a.

----- 1978. Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs, Volume III: Year Two Impact Designs. American Institutes for Research. Palo Alto.

----- 1978b. Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Programs, Volume IV: Overview of the Study and Findings. American Institutes for Research. Palo Alto.

Educational Operations Concepts, Inc. 1991b. An Evaluation of the Title VII ESEA Bilingual Education Program for Hmong and Cambodian Students in Kindergarden and First Grade St. Paul.  
Redundant with Educational Operations Concepts, Inc 1991a.

El Paso Independent School District. 1992. Bilingual Education Evaluation Office for Research and Evaluation. El Paso.  
Redundant with El Paso 1987.

El Paso Independent School District. 1990. Bilingual Education Evaluation: The Sixth Year in a Longitudinal Study. Office for Research and Evaluation. El Paso.

Genesee, Fred, Lambert, Wallace E., and Tucker, G.E. 1977. An Experiment in Trilingual Education. McGill University. Montreal.  
Redundant with Genesee et al 1983.

McConnell, Beverly Brown. 1980b. "Individualized Bilingual Instruction, Final Evaluation, 1978-1979 Program." Pullman.

Redundant with McConnell 1980a.

----- 1980c. "Individualized Bilingual Instruction for Migrants." Paper presented at the October meeting of the International Congress for Individualized Instruction. Windsor.

McSpadden, J.R.. 1980. Arcadia Bilingual Bicultural Education Program: Interim Evaluation Report, 1979-80. Lafayette Parish.  
Redundant with McSpadden 1979.

Teschner, Ricard V. 1990. "Adequate Motivation and Bilingual Education." Southwest Journal of Instruction. Volume 9. pp. 1-42.  
Redundant with El Paso, 1990.

#### Studies Excluded Because They are Unavailable

American Institutes for Research. 1975b. "Bilingual Education Program (Aprendamos En Dos Idiomas). Corpus Christi. Identification and Description of Exemplary Bilingual Education Programs. Palo Alto.

Lambert, Wallace E., and Tucker, G. R. 1972. Bilingual Education of Children: The St. Lambert Experience. Newbury House. Rowley.

McSpadden, J.R. 1979. Arcadia Bilingual Bicultural Education Program: Interim Evaluation Report, 1978-79. Lafayette Parish.

Morgan, Judith Claire. 1971. "The Effects of Bilingual Instruction of the English Language Arts Achievement of First Grade Children." Ph.D. dissertation. Northwestern State University of Louisiana.

Ramos, M., Aguilar, J.V., and Sibayan, B.F. 1967. The Determination and Implementation of Language Policy. Phillipine Center for Language Study: Monograph Series 2. Quezon City.

#### Studies Excluded Because They are not Evaluations of Bilingual Programs

Becker, Wesley C. and Gersten, Russell. 1982. "A Follow-Up of Follow Through: The Latter Effects of the Direct Instruction Model on Children in Fifth and Sixth Grades." American Educational Research Journal. Vol. 19. pp. 75-92.

Campeau, Peggie L., Roberts, A. Oscar H., Bowers, John E., Austin, Melanie, and Roberts, Sarah J. 1975. The Identification and Description of Exemplary Bilingual Education Programs. American Institutes for Research. Palo Alto.

Webb, John A., Clerc, R.J., and Gavito, Alfredo. 1987. Houston Independent School District: Comparison of Bilingual and Immersion Programs Using Structural Modeling. Houston Independent School District.

#### Studies Excluded Because There is not an Appropriate Control Group

Barik, Henri, Swain, Merrill. and Nwanunobi, E. A. 1977. "English-French Bilingual Education: The Elgin Study Through Grade Five." Canadian Modern Language Review. Vol. 33. pp. 459-475.

Bruck, Margaret, Lambert, Wallace E., and Tucker, G. Richard. 1977. "Cognitive Consequences of Bilingual Schooling: The St. Lambert Project Through Grade Six." Linguistics. Vol. 24. pp. 13-33.

Burkheimer, Graham J., Conger, A.J., Dunteman, G.H., Elliott, B.G., and Mowbray, K.A. 1989. Effectiveness of Services for Language-Minority Limited-English-Proficient Students. Report to the U.S. Department of Education.

Day, Elaine M., and Shapson, Stan M. 1988. "Provincial Assessment of Early and Late French Immersion Programs in British Columbia, Canada." Paper presented at the April meetings of the American Educational Research Associates. New Orleans.  
No background controls or individual level data reported.

El Paso Independent School District. 1987. Interim Report of the Five-Year Bilingual Education Pilot 1986-1987 School Year. Office for Research and Evaluation. El Paso.  
No background or pretest controls.

Genesee, Fred., and Lambert, W. E. 1983. "Trilingual Education for Majority-Language Children." Child Development. Volume. 54. pp. 105-114.  
No background controls.

Genesee, Fred, Holobow, Naomi E., Lambert, Wallace E, and Chartrand, Louise. 1989. "Three Elementary School Alternatives for Learning Through a Second Language. The Modern Language Journal. Volume. 73. pp. 250-263.  
No background controls.

Gersten, Russell. 1985. "Structured Immersion for Language-Minority Students: Results of a Longitudinal Evaluation." Educational Evaluation and Policy Analysis. Volume. 7. pp. 187-196.  
No background controls.

Malherbe, E. C. 1946. The Bilingual School. Longmans Green. London.  
No background or pretest controls.

McConnell, Beverly Brown. 1980a. "Effectiveness of Individualized Bilingual Instruction for Migrant Students." Ph.D. dissertation. Washington State University

Medina, Marcello, and Escamilla, Kathy. 1992. "Evaluation of Transitional and Maintenance Bilingual Programs." Urban Education. Volume 27. No. 3. p. 263-290.

Melendez, William Anselmo. 1980. "The Effect of the Language of Instruction on the Reading Achievement of Limited English Speakers in Secondary Schools." Ph.D. dissertation. Loyola University of Chicago.  
No background controls.

Stern, Carolyn. 1975. Final Report to the Compton Unified School District's Title VII Bilingual/Bicultural Project: September 1969 Through June 1975. Compton City Schools. Compton.

Vasquez, Miriam. 1990. "A Longitudinal Study of Cohort Academic Success and Bilingual Education." Ph.D. dissertation. University of Rochester.  
No background controls.

Studies Excluded Because the Effects are Measured after an Unreasonably Short Period

Barclay, Lisa. 1969. "The Comparative Efficacies of Spanish, English, and Bilingual Cognitive Verbal Instruction with Mexican American Head Start Children." Ph.D. dissertation. Stanford University.  
Positive Average Effect.

Layden, Russell Glenn 1972. "The Relationship Between the Language of Instruction and the Development of Self-Concept, Classroom Climate, and Achievement of Spanish Speaking Puerto Rican Children." Ph.D. dissertation. University of Maryland.  
Negative Average Effect.

Studies Excluded Because They Inadequately Control for Differences between Bilingual and English-Only Students

Alvarez, Juan. 1975. "Comparison of Academic Aspirations and Achievement in Bilingual Versus Monolingual Classrooms." Ph.D. dissertation. UT Austin.  
Negative Average Effect.

Ames, J., and Bicks, Pat. 1978. An Evaluation of Title VII Bilingual/Bicultural Program, 1977-1978 School Year, Final Report. Community School District 22. Brooklyn. School District of New York.  
Positive Average Effect.

Balasubramonian, K., Seelye, H., and Elizondo de Weffer, R. 1973. "Do Bilingual Education Programs Inhibit English Language Achievement: A Report on An Illinois Experiment." Paper presented at the 7th Annual Convention of Teachers of English to Speakers of Other Languages. San Juan.  
Positive Average Effect.

Barik, Henri, and Swain, Merrill. 1975. "Three Year Evaluation of a Large-Scale Early Grade French Immersion Program: The Ottawa-Study." Language Learning. Vol. 25. No. 1. pp. 1-30.  
Negative Average Effect.

Bates, Enid May Buswell. 1970. "The Effects of One Experimental Bilingual Program on Verbal Ability and Vocabulary of First Grade Pupils." Ph.D. dissertation. Texas Tech University.  
Negative Average Effect.

Carsrud, Karen, and Curtis, John. 1980. ESEA Title VII Bilingual Program: Final Report. Austin Independent School District. Austin.  
No statistical tests reported.  
Positive Average Effect.

Ciriza, Frank. 1990a. Evaluation Report of the Preschool Project for Spanish-Speaking Children, 1989-1990. Planning, Research and Evaluation Division. San Diego City Schools. San Diego.  
Positive Average Effect.

Cohen, Andrew D. 1975. A Sociolinguistic Approach to Bilingual Education. Newbury House Press. Rowley, MA.  
Negative Average Effect.

Cottrell, Milford C. 1971. "Bilingual Education in San Juan Co., Utah: A Cross-Cultural Emphasis." Paper presented at the April meetings of the American Educational Research Association. New York City.  
Negative Average Effect.

Curiel, Herman. 1979. "A Comparative Study Investigating Achieved Reading Level, Self-Esteem, and Achieved Grade Point Average Given Varying Participation." Ph. D. dissertation. Texas A&M.  
Negative Average Effect.

de Weffer, Rafalea de Carmen Elizondo. 1972. "Effects of First Language Instruction in Academic and Psychological Development of Bilingual Children." Ph.D. dissertation. Illinois Institute of Technology.  
Positive Average Effect.

de la Garza, Jesus Valenzuela, and Marcella, Medina. 1985. "Academic Achievement as Influenced by Bilingual Instruction for Spanish-Dominant Mexican American Children." Hispanic Journal of Behavioral Sciences. Volume. 7. No. 3. pp. 247-259.  
Positive Average Effect.

Educational Operations Concepts, Inc. 1991a. An Evaluation of the Title VII ESEA Bilingual Education Program for Hmong and Cambodian Students in Junior and Senior High School. St. Paul.  
Positive Average Effect.

Lampman, Henry P. 1973. "Southeastern New Mexico Bilingual Program: Final Report." Artesia Public Schools. Artesia.  
Positive Average Effect.

Legarreta, Dorothy. 1979. "The Effects of Program Models on Language Acquisition by Spanish-Speaking Children." TESOL Quarterly. Volume. 13. No. 4. pp. 521-534.  
Positive Average Effect.

Lum, John Bernard. 1971. "An Effectiveness Study of English as a Second Language (ESL) and Chinese Bilingual Methods." Ph.D. dissertation. U.C. Berkeley  
Negative Average Effect.

Maldonado, Jesus Ruben. 1974. "The Effect of the ESEA Title VII Program on the Cognitive Development of Mexican American Students." Ph.D. dissertation. University of Houston.  
Negative Average Effect.

Matthews, T. 1979. "An Investigation of the Effects of Background Characteristics and Special Language Services on the Reading Achievement and English Fluency of Bilingual Students." Seattle Public Schools: Department of Planning, Research, and Evaluation. Seattle.  
Negative Average Effect.

Moore, Fernie B. and Parr, Gerald D. 1978. "Models of Bilingual Education: Comparisons of Effectiveness." The Elementary School Journal. Volume. 79. pp. 93-97.  
Negative Average Effect.

Pena-Hughes, Eva, and Solis, Juan. 1980. ABC's. McAllen Independent School, District. McAllen.  
Positive Average Effect.

Prewitt Diaz, Joseph O. 1979. "An Analysis of the Effects of a Bicultural Curriculum on Monolingual Spanish Ninth Graders as Compared with Monolingual English and Bilingual Ninth Graders with Regard to Language Development, Attitude Toward School, and Self-Concept." Ph.D. dissertation. University of Connecticut.  
Positive Average Effect.

Stebbins, Linda B., St. Pierre, Robert G., Proper, Elizabeth C., Anderson, Richard B., and Carva, Thomas. 1977. "Education as Experimentation: A Planned Variation Model, Volume IV-A. An Evaluation of Follow Through." ABT Associates. Cambridge.  
Positive Average Effect.

Valladolid, Lupe A. 1991. "The Effects of Bilingual Education of Students' Academic Achievement as They Progress Through a Bilingual Program." Ph.D. dissertation. United States International University.  
No background or pretest controls.  
Negative Average Effect.

Yap, Kim O., Enoki, Donald Y., and Ishitani, Patricia. 1988. "SLEP Student Achievement: Some Pertinent Variables and Policy Implications." Paper presented at the April meetings of the American Educational Research Association. New Orleans.  
No background or pretest controls.  
Negative Average Effect.

Zirkel, Perry A. 1972. "An Evaluation of the Effectiveness of Selected Experimental Bilingual Education Programs in Connecticut." Ph.D. dissertation. University of Connecticut.  
Positive Average Effect.

#### Other Sources

Baker, Keith. 1987. "Comment on Willig's 'A Meta-Analysis of Selected Studies in the Effectiveness of Bilingual Education.'" Review of Educational Research. Volume 57, pp. 351-362.

Baker, K.A. and de Kanter, A.A. 1981. Effectiveness of bilingual education: A review of the literature. Washington, D.C.: U.S. Department of Education, Office of Planning, Budget and Evaluation.

Campbell, D. T. and Erlebacher, A. E. 1970. How regression artifacts in quasi-experimental evaluations can mistakenly make compensatory education look harmful. In J. Hellmuth (Ed.) Compensatory Education: A National Debate (Volume 3). Disadvantaged Child. New York: Brunner/Mazel.

- Cook, T. D., et al. 1992. Meta-Analysis for Explanation: A Casebook. New York: Russell Sage Foundation.
- Greene, J.P., Peterson, P.E., and Du, J. 1997. Effectiveness of School Choice: The Milwaukee Experiment. Harvard Program on Education Policy and Governance Working Paper 97-1.
- Hanushek, Eric A. 1996. "School Resources and Student Performance." In Does Money Matter, ed. Gary Burtless. Washington, D.C.: Brookings, pp. 43-73.
- Hedges, Larry V. and Rob Greenwald. 1996. "Have Times Changed?" .” In Does Money Matter, ed. Gary Burtless. Washington, D.C.: Brookings, pp. 74-92.
- Krashen, S. 1996. Under Attack: The Case Against Bilingual Education. Culver City, CA: Language Education Associates.
- Rosenthal, R. 1991. Meta-Analytic Procedures for Social Research. Newbury Park: Sage Publications.
- Rossell C.H. and Baker K. 1996. "The Educational Effectiveness of Bilingual Education." Research in the Teaching of English, Vol. 30, no 1.
- Willig, A. 1985. "A Meta-Analysis of Selected Studies on the Effectiveness of Bilingual Education," Review of Educational Research, volume 55, no. 3.
- Willig, A. 1987. "Examining Bilingual Education Research Through Meta-Analysis and Narrative Review: A Response to Baker." Review of Educational Research, vol 57, no 3.