DOCUMENT RESUME

ED 283 932 UD 025 648

AUTHOR	Rossell, Christine H.
TITLE	Why Is Bilingual Education Research So Bad? A Critique of the Walsh and Carballo Study of Massachusetts Bilingual Education Programs. Working Paper 86-5.
INSTITUTION	Boston Univ., MA. Center for Applied Social Science.
PUB DATE	Oct 86
NOTE	22p.
PUB TYPE	Information Analyses (070) -- Viewpoints (120)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Bilingual Education Programs; Control Groups; *English Instruction; *Experimental Groups; Limited English Speaking; Reliability; *Research Methodology; *Research Problems; Sample Size; Statistical Analysis; *Transitional Programs
IDENTIFIERS	Evaluation Research; *Massachusetts

ABSTRACT

The Walsh and Carballo evaluation of the effectiveness of transitional bilingual education programs in five Massachusetts communities, has the following flaws: (1) the sample of school districts studied suffered from "self-selection bias"; (2) the sample does not include a single large, urban school district; (3) the student samples analyzed are much too small to allow for any conclusions; (4) there is no statistical analysis of the data nor control for pre-existing differences between groups; and (5) the wrong comparison is conducted. In elaborating on those flaws, the report first describes five models of how to instruct children who do not speak English. It then summarizes the requirements of a methodological sound study and maintains that the studies cited by Walsh and Carballo do not follow those rules. Finally, the above five flaws are discussed in detail. A bibliography is included. (PS)

7-2-87

WHY IS BILINGUAL EDUCATION RESEARCH SO BAD? A CRITIQUE OF THE WALSH AND CARBALLO STUDY OF MASSACHUSETTS BILINGUAL EDUCATION PROGRAMS

Christine H. Rossell

CENTER FOR APPLIED SOCIAL SCIENCE

Boston University
Boston, Massachusetts 02215

WHY IS BILINGUAL EDUCATION RESEARCH SO BAD? A CRITIQUE OF THE WALSH AND CARBALLO STUDY OF MASSACHUSETTS BILINGUAL EDUCATION PROGRAMS

Christine H. Rossell

Center for Applied Social Science
BOSTON UNIVERSITY
232 Bay State Road
Boston, MA 02215

WORKING PAPER
86-5

October, 1986

Christine H. Rossell, Ph.D.: Research Associate of the Center for Applied Social Science; Associate Professor of Political Science

WHY IS BILINGUAL EDUCATION RESEARCH SO BAD? A CRITIQUE OF THE WALSH AND CARBALLO STUDY OF MASSACHUSETTS BILINGUAL EDUCATION PROGRAMS

Introduction

If there is one aspect of the research on bilingual education that all commentators agree on, it is that its quality is deplorable. One major problem is that most of it consists of local evaluations with inadequate research designs and analyses. These local evaluators are usually unable or unwilling to assemble a comparison group of students who have not had bilingual education, and typically assess only gains for the students in bilingual education before and after their participation in the program. To their credit, many evaluators forced to use this model understand they can draw no policy conclusions from it. Unfortunately, many do not, and numerous reviewers have compounded the error by uncritically citing these and other flawed studies as support for transitional bilingual education as the best policy alternative for producing the greatest English language achievement in children of limited English language proficiency (LEP).

The second characteristic of the research is that, as is common with controversial social programs with egalitarian goals, the evaluators and those who review and integrate the research are also passionate advocates of bilingual education for political or ideological reasons. The disgraceful treatment of linguistic minorities in this country -- the mislabeling of limited English proficiency children as mentally retarded, their high dropout

or pushout rate because they have been allowed to flounder in an alien, hostile environment or actually punished for using their mother tongue -- may have influenced many social scientists, bilingual education lawyers, and reviewers of the research to believe that any policy which ignores the mother tongue in favor of English is racist, and any policy which maintains the mother tongue, however inadequately, is equitable. This has created an atmosphere in which it is difficult for an academic to criticize current policy in this field. It has also created an atmosphere in which it is all too easy to interpret flawed studies as support for bilingual education and to reject or ignore competent, relevant studies with conflicting findings.

A recent evaluation of transitional bilingual education programs in five Massachusetts communities by Catherine E. Walsh and Eduardo B. Carballo, Transitional Bilingual Education in Massachusetts: A Preliminary Study of Its Effectiveness (April 1986), follows in this unfortunate tradition. The authors misrepresent the previous research in the field, perform an evaluation so inadequate it can tell us nothing about the effectiveness of any of the bilingual education programs they studied, and then rather than apologizing, proclaim their study as evidence of the success of transitional bilingual education. Specifically, the Walsh and Carballo evaluation has the following problems:

1) The sample of five school districts suffers from "self-selection bias." Only school districts willing to be evaluated, presumably those with the most effective programs, were included in the study.

2) Probably as a result of their selection criterion, the sample does not include a single large, urban school district. Most glaring is the omission of Boston and Springfield.

3) The student samples analyzed are much too small to draw any conclusions from. Most problematic is the fact that the control

4) comparison group for each of the five school districts is either grossly inadequate or just plain nonexistent.

4) There is no statistical analysis of the data nor control for pre-existing differences between groups.

5) Walsh and Carballo conduct the wrong comparison. Students in transitional bilingual education are compared to students who have received no help at all rather than students in alternative programs.

What is Bilingual Education?

There are five models of how to instruct children who do not speak English. The first is the old "sink-or-swim" method which was made illegal as educational policy by Lau v. Nichols, 1974. Here students are simply placed in the regular classroom with no special help regardless of their English language ability. No responsible educator or social scientist advocates such a policy since there are more humane and effective models.

A second instructional technique is English as a Second Language (ESL) instruction for one or two periods a day, and "submersion" in the regular classroom for the rest of the day. ESL is a pull-out program usually based on a special curriculum, but the instructors do not have to know the child's native language.

A third policy alternative is structured immersion where instruction is in the language being learned (L2), but the teacher knows the students' native tongue (L1). The L2 (i.e. English) used in these programs is always geared to the children's language proficiency at each stage so that it is comprehensible. L1 is used only in the rare instances when the student cannot complete a task without it. The student thus learns L2 and subject matter content simultaneously. Immersion programs in which L2 is not the dominant language of the country typically include at least 30-60 minutes a day of L1 language arts. In fact, most of the Canadian "immersion" programs,

where English-speaking students learn French, eventually become bilingual programs.

A fourth, and the most widely implemented, policy alternative is transitional bilingual education (TBE). According to Young, et al. (1984) at least 40 percent of all limited English proficient (LEP) children are now in TBE programs, and only 2.6 percent are in English instruction classrooms. The other 34 percent are divided among bilingual maintenance, Spanish instruction, and ESL classes. By contrast, Okata, et al. (1983) found no projects which reported "English only" as a literacy goal for LEP students. Hence, TBE is clearly the dominant special language instructional program in the U.S.

In transitional bilingual education, the student is taught both in the native tongue (L1) and the language being learned (L2), with subject matter taught in L1. The amount of instructional time in L1 is reduced, and L2 increased, until the student is proficient enough in L2 to join the regular instructional program. The majority of elementary school programs are three year programs. The rationale underlying TBE differs depending on the age of the child. For very young children, it is that learning to read in the native tongue first is a necessary condition for optimal reading ability in the second tongue. For all children, it is argued that learning a second language takes time and children should not lose ground in other subject matters, particularly math, during the time period they are learning the second language.

A final model is bilingual maintenance. Rather than being transitioned out of bilingual education, students remain in the program for their entire school career. The goals of this model are social and intellectual rather

than remedial and therapeutic as with the other models. Almost all such programs in the United States are bilingual education magnet programs which are racially and ethnically integrated in order to desegregate a school system.

The Walsh and Carballo study is an evaluation of transitional bilingual education programs in five Massachusetts communities, Attleboro, Cambridge, Framingham, Haverhill and Holyoke. Their study compared students in transitional bilingual education (TBE) to 1) students who have graduated from TBE programs and have been "mainstreamed" into a regular classroom and 2) a control group of students who have received no, or "minimal," services, i.e. submersion.

Previous Research

Walsh and Carballo begin their evaluation of Massachusetts programs by reviewing some of the research in the field. They criticize national studies which have shown transitional bilingual education to be ineffective and praise studies that purport to demonstrate the effectiveness of transitional bilingual education. They do this with no regard for the methodological standards of social science research.

What is a methodologically sound study? In order to determine whether a bilingual education program is successful, the research study must have a treatment group subjected to the program and a control or comparison group, similar to the treatment group, which has not received that program. If students have not been randomly assigned to the control group, there must be statistical control for differences between the groups which existed prior to the time one group received bilingual education. Post-bilingual education differences between groups alleged to have been caused by the bilingual

education program must be tested by means of an appropriate statistical analysis to determine if the differences are greater than could have been expected by chance controlling for pre-existing differences.

The reason why it is essential to have a comparison group is because all children tend to show progress in English language knowledge over time. Children of limited-English-proficiency will know more English the longer they are in any program regardless of its effectiveness. Unless we have a comparison group not receiving that program, we will not know if that increased achievement is more or less than we would expect to occur naturally. For example, if a child enters a transitional bilingual education program with an English language score of 20 and comes out with a score of 60 that might actually be a negative program effect if similar children in other programs, or in no program at all, are scoring 20 when they enter school and 80 after the same time period. Hence, a comparison group is absolutely essential to program evaluation. Nevertheless, it is missing from most evaluations of bilingual education because of the difficulty of finding similar children of limited English proficiency who are not in a bilingual education program. Rather than apologizing for the lack of a control group, and suggesting that as a result no conclusions can be drawn, all too many evaluations conclude that because children know more English after participation in the TBE program than before, the program is a success.

Even if there is a comparison group, pre-existing differences between groups must be statistically controlled for. This is because children with higher achievement prior to bilingual education will tend to have higher achievement after bilingual education even if the bilingual education lowered their English language achievement. In addition, children of higher

socioeconomic status will tend to have higher achievement after bilingual education than children of lower socioeconomic status not in the program even if the TBE program lowered the achievement of the higher socioeconomic students.

Not one of the studies cited by Walsh and Carballo as evidence of the effectiveness of transitional bilingual education followed these absolutely essential rules for determining program effect nor did the reviews they cited select only studies which did that.[1] Ironically, the two studies dismissed by Walsh and Carballo as "methodologically at fault" did follow these essential rules for determining program effect. The AIR study (Danoff, et al., 1977; 1978) not only had a control group, but controlled for pre-existing differences between the students in bilingual education and those in the control group. The Baker and de Kanter review (1981) selected only studies that had these characteristics. Both the AIR and Baker and de Kanter studies also had larger samples than any other study or review to date. They both found transitional bilingual education to be ineffective in comparison to other programs. Hence, one suspects that what Walsh and Carballo really object to are the findings of these two studies, not their methodology.

Of the methodologically sound studies I reviewed (Rossell and Ross, 1986), 71 percent found transitional bilingual education to be no different or worse than doing nothing in second language learning and 93 percent showed

---

[1] For a detailed critique of the studies they cited as evidence of the success of bilingual education, see Christine H. Rossell and J. Michael Ross, "The Social Science Evidence on Bilingual Education," Center for Applied Social Science Working Paper 85-7, Boston University or the same article in The Journal of Law and Education, Fall 1986; and Keith A. Baker and Adriana de Kanter, Effectiveness of Bilingual Education: A Review of the Literature, 1981. Both reviews clearly delineate the standards for methodologically acceptable studies and the shortcomings of most studies in the field.

it to be no different or worse than doing nothing in math learning. All but one comparison of transitional bilingual education to structured immersion showed the latter to be superior in both second language and math learning.

Thus, Walsh and Carballo begin their evaluation with an inaccurate and misleading review of the research in this field.

Unfortunately, their conclusions regarding the consensus in this field are all too accurate. The field of bilingual education is pervaded by a disregard for the canons of scientific research.

School District Sample

There are two problems with the sample of five school districts analyzed in the Walsh and Carballo study. First, the sample suffers from "self-selection bias." The only school districts that were studied were those that agreed to be studied.[2] Selecting only school districts willing to be studied is unacceptable by social science research standards because the school districts which refuse to participate are likely to be those with unsuccessful programs.

Second, there is not a single large, urban school district in the sample. This is probably a result of Walsh and Carballo's selection criterion - that is, only school districts willing to participate were studied. The lack of a large, urban school district is important because it is easier to implement any program on a small scale even if the program itself is not a particularly

---

[2] It is not clear why this limitation was placed on the study since one of the co-authors, Eduardo Carballo, is an administrator in the Massachusetts Department of Education and presumably could have insisted on all school districts participating.

good one. Thus, the generalizability of any findings from this study will be limited. The findings cannot be applied to the two cities, Boston and Springfield, with the largest number of limited English proficient students in the state.

Student Sample and Control Group

The sample size for two of the three groups of students Walsh and Carballo studied is so inadequate as to disqualify the study on these grounds alone. To reiterate, the three groups of students studied are: 1) students in transitional bilingual education (TBE), 2) students formerly in bilingual education but now "mainstreamed" into regular classrooms, and 3) a control group of students identified by administrators as limited English proficient but whose parents refused to enroll them in bilingual education.

The sample size in Attleboro is 16 TBE students, 11 mainstreamed students, and 0 control group students. The sample size in Cambridge is 25 TBE students, 5 mainstreamed students, and 11 control group students. The sample size in Framingham is 27 TBE students, 18 mainstreamed students, and 3 control group students. The sample size in Haverhill is 18 TBE students, 18 mainstreamed students, and 8 control group students. The sample size in Holyoke is 43 TBE students, 7 mainstreamed students and 5 control group students. Since all comparisons are by school district, the mainstream and control group are completely inadequate before any measures of program success, with their accompanying missing data, are analyzed.

Sample Size and Its Effect on Measures of Program Success

Attendance. On the first measure of program success, days of attendance, we are given no information as to the size of the sample in each category or cell.[3] Again, this is unacceptable by social science standards. Moreover, TBE students and mainstreamed students are inexplicably collapsed into one group so that we cannot tell how many mainstreamed students are in each category of attendance -- 180-160 days, 160-140 days, 140-120 days, and 120 or less. In addition, since there is no statistical control for their measure of social class (whether a student is receiving free or reduced lunch), we also have no idea as to what extent the observed attendance rates are explained by the social class of each group.

Sixty-eight percent of the TBE/mainstream group in Attleboro attends school for 180-160 days a year, but since there is no control group in this district (and no data on almost 20 percent of the TBE/mainstream sample), we have no idea whether this is better or worse than the rest of the students in that school district.

---

[3] This is true of all tables. The reader never knows how many students the authors are analyzing in each category. The exception to this is when there are 0 or only 1 student in a cell. This is indicated by an asterisk.

[4] Walsh and Carballo give us no information on the social class composition of each of the three groups: TBE, mainstreamed, and control group. It is thus possible, although we have no way of telling, that the control group has much lower social class than the TBE or mainstreamed groups. This is one of the many glaring errors which render this study unintelligible. Moreover, they only collected social class data on 56 percent of the sample in Attleboro, 68 percent of the population in Cambridge, 61 percent of the sample in Haverhill, and 75 percent of the sample in Holyoke. Yet, they were able to obtain social class data on 91 percent of the population in Framingham. Thus, for four out of five school districts, there is an unacceptably high rate of missing data on this variable, particularly given the small sample size.

In Cambridge, 70 percent of the TBE/mainstream students attend school 180-160 days a year, but since 80 percent of the control group attend school 180-160 days, this seemingly positive attendance rate is actually a negative program effect.[5] In Framingham, 65 percent of the TBE/mainstream students attend school 180-160 days, but since 100 percent of the control group attends school 180-160 days, this seemingly positive program effect is actually a negative effect.[6] Only in Haverhill and Holyoke do TBE/mainstream students have a higher rate of attendance than control group students, but given the small size of the control group and the expected missing data, we may again be talking about only a couple of students. Although for all the reasons mentioned we can draw no conclusions from it, the Walsh and Carballo data actually show that in 50 percent of the school districts for which there is a control group, the TBE/mainstream students had higher attendance and in 50 percent they had lower attendance than the control group.

Walsh and Carballo may have concluded TBE had a positive effect on attendance from their "totals percentages" for the four school districts with a control group. These "totals percentages" show 70 percent of the TBE/mainstream students attending 180-160 days but only 58 percent of the

---

[5] This finding demonstrates how important a control group is in evaluating programs. Without a control group, we would have concluded the program had a positive effect on children's attendance. With a control group, we can see the TBE program actually had a negative effect. Moreover, Cambridge is the only school district on which we are given information (in the narrative) about the social class of the TBE group and the control group. The control group is of lower class than the TBE group. Whereas 71 percent of the TBE students are on free or reduced lunch, 86 percent of the control group are on free or reduced lunch. Thus, despite the fact that they were of lower social class, the control group had higher attendance.

[6] Of course, the Framingham control group of 3 students (or less) is too small to draw any conclusions from, but Walsh and Carballo do not seem to know this.

control group students attending 180-160 days a year. Unfortunately, their "total percent" does not appear to be a total, but an average of the percentages for each school district. The problem with an average is that it weights each school district equally rather than by their sample size. Given my suspicion that Cambridge is not the only school district where the control group is of lower social class than the TBE/mainstream students, a statistical analysis controlling for the social class differences between the two groups may find the control group to be superior in school attendance.

Grades. Grades are an unreliable source of program success when comparing TBE students to control group students. Students in a class are always graded in comparison to other students in their class not to students in other programs. Even if none of the students in a class know very much English, some of them will still receive high grades. Thus the average grade of B for students in TBE tells us nothing about how much English they know or are learning.

The grades of mainstream students are another matter. They are competing against English-speaking students in the regular classroom. Although there will still be some tendency on the part of teachers toward grade inflation, it should be small unless a school or school district practices academic tracking. Thus, if the sample size of the mainstream and control group were adequate, we might actually have a real possibility of assessing program success for these groups with this variable. Unfortunately, both the mainstream and control group student sample ranges from 0 to 11 before analysis of their grades is conducted and missing data is taken into account. Of the 20 cells in the mainstreamed and control groups (Table 6, page 44), 30 percent are empty or have only one student. The rest could have only

two students in each cell. It is impossible to do any valid comparisons with such minuscule sample sizes.

English Language Achievement. Clearly, the most important measure of program success is English language achievement. The purpose of special language programs for limited English proficient students is to teach them English. It is of very little importance if they attend school regularly and have high grades in comparison to other limited English proficient students if they do not know English. Walsh and Carballo, however, do not consider this an important enough variable to present in a table as with the other measures of program success. One has to wade through the narrative on each school district to determine the effect of transitional bilingual education on English language achievement.

Unfortunately, although this is the most important variable, the least amount of data was collected on it. Of the five school districts with a control group in this sample, three (Attleboro, Cambridge, and Holyoke) have no achievement data whatsoever for the control group students. A fourth school district (Framingham) has one control group achievement score, but none for the TBE students. The fifth school district, Haverhill, has 6 control group student achievement scores but only one TBE student achievement score. There are, however, achievement scores for 12 mainstreamed students. They are doing very poorly. In short, there is not enough data on achievement to conduct even the crudest of comparisons, let along the correct one which would statistically compare the groups and control for pre-existing differences.

As Walsh and Carballo admit, elementary TBE students average one year below grade level equivalencies and secondary TBE students two years below grade level equivalencies. What they do not acknowledge is that without a

control group we have no idea whether that is a positive program effect, a negative one, or no effect at all.

Statistical Analysis

As stated several times in this paper social science research has rules for determining whether one has proven one's hypothesis. Walsh and Carballo should have statistically compared TBE students to control group students, or students in alternative programs, and mainstreamed students to control group students, or students in alternative programs, to see if they differed more than would have been expected by chance given their sample size and variance within each group controlling for the social class of the students and their pre-program English language achievement. Walsh and Carballo conducted not a single statistical comparison because, given their minuscule sample size, they could not. The only question remaining is why present any data at all if it is insufficient to be analyzed by valid social science methods to determine the effectiveness of the program?

The Wrong Comparison

Even had they had an adequate sample and correctly analyzed the differences between groups, the Walsh and Carballo study could still have been criticized for comparing TBE and mainstreamed students to the wrong group. The control group in this study consists of students who received no, or "minimal," services[7]. I know of no educator or social scientist who

[7] Walsh and Carballo state that "although control group students are enrolled in the monolingual curriculum, those identified as LEP (limited English proficient) are monitored by the TBE program and, depending on school location, are offered minimal ESL support" (fn. 4). One cannot help but have the suspicion that administrators were able to identify the control

would advocate no, or minimal, services for LEP children. Children of limited English proficiency need special help and should not be left to sink or swim in the regular classroom when there are more effective and humane alternatives.

The only policy proposals that I have seen recommend alternative forms of special language help for limited English proficient children. One very successful alternative program is structured immersion, described on page 4. Moreover, there are some programs similar to structured immersion in Newton and Brookline.[8] The students in these programs could be compared to those in transitional bilingual education controlling for social class and pre-existing English language ability. In addition, transitional bilingual education programs vary in the extent of English used in the classroom. Programs which use more English could be compared to those which use less English. Many students are placed in ESL programs because the school

---

group students because they were having academic difficulty and thus were being monitored. It is hard to believe, particularly given the high rate of missing data in this study that administrators are able to keep track of every single student who entered the school system, was identified as LEP, but whose parents did not want him or her enrolled in transitional bilingual education or for whom there was no program. If they were able to do that, the control group in this sample would have been much larger and included numerous Asian linguistic minorities for whom it is impossible to conduct a TBE program because there are no certified teachers who speak their language. Given the unlikeliness of identifying every single formerly LEP student, administrators would naturally tend to identify as LEP students those who are currently LEP, that is having academic difficulties. Thus, there might also be a student selection bias here which needs to be controlled for, although unfortunately no statistical technique will completely do that.

[8] Despite the fact that the research shows structured immersion to be superior to transitional bilingual education and transitional bilingual education to be no better than doing nothing, the administrators of these programs find themselves to be in violation of state law. Even more administrators would be willing to adopt similar alternative models, but lack the courage or the funds to fight state agencies.

district cannot find a certified teacher who speaks their language or there are not enough students with that language to justify allocating a classroom and teacher. Cambridge, for example, has some students in ESL programs with academic content area instruction conducted in regular classrooms. These students could be compared to TBE students controlling for pre-program differences between the groups.

None of these data have been collected by Walsh and Carballo or any other evaluators. However, it is exactly this kind of information on variations in TBE and alternative programs which is needed to assemble a respectable "control" or comparison group. The only mystery is why it hasn't been done.

Conclusions

The Walsh and Carballo study of transitional bilingual education programs in five Massachusetts communities does not show that "TBE students and mainstream students are much more successful in school than are LEP students who have never been served by TBE programs (control students)" as the authors claim (p.73). Even if we are to take the data at face value and not demand that it conform to the standards of social science research, it does not show that. Days of attendance is not a measure of success in school, but even if it were, half the cases show TBE to be superior and half show doing nothing to be better. Grades are a better measure of success in school, but there is virtually no data for the mainstreamed and control group students and so one can come to no conclusion regarding success in school when limited English proficient students are compared only to each other.

The best measure of what transitional bilingual education is supposed to be accomplishing is a measure of English language achievement, but there is also virtually no comparative data on that.[9] Either the TBE group or the control group are missing achievement data in all but one of the school districts. In the one school district were there is achievement data for the control group and the mainstreamed students, the sample size is so small (6) that an analysis of covariance could not be conducted even if Walsh and Carballo wished to do so.

It would be nice if I could say this is one of the worst evaluations of bilingual education I have ever read, but that is not the case. The research in this field is so bad that this study probably ranks in the top half of all evaluations in terms of quality. Most local evaluations do not even attempt to assemble a control group nor examine the progress of mainstreamed students as Walsh and Carballo did. In a sense then, they are more clever than most evaluators since they appear to have comparison groups, but ultimately do not.

An important question usually ignored in discussions about the poor quality of bilingual education research (Zappert and Cruz, 1977; Okata, 1983; Willig, 1981-82), is why is it so bad and why is this tolerated? I believe the research is poor because bilingual education is the ideologically "correct" policy alternative. To be in favor of bilingual education, regardless of its effect on children, is the "civil rights" position. To be in favor of alternatives to

---

[9] No explanation is given as to me why Walsh and Carballo had such a hard time collecting data on their outcome measures. School districts routinely test all students yearly and new students at the time of admission. They also keep track of the attendance of all students and their grades. Yet this routine bookkeeping information is missing for more than half of their already minuscule sample.

bilingual education is to be reactionary and racist. The reason this has occurred is because no other policy alternative allows use of the native tongue in instruction and also requires the use of native tongue speakers as instructors. Thus, in the minds of many civil rights advocates, this feature of bilingual education is so important as to make its effect on English language achievement secondary in importance. Nevertheless, it is obviously politically useful to show a positive English language achievement effect. Since a poor evaluation -- that is, one with no comparison group and/or statistical analysis -- will guarantee a "positive" English language achievement effect, all but a handful of bilingual education evaluations are of poor quality.

The elites, academics, and policymakers who tolerate such poor research generally fall into two groups: those who have been intimidated by the bilingual education "establishment" into supporting TBE and those who simply do not care enough about the education of immigrant children to determine the truth for themselves. Perhaps one of the saddest aspects of the bilingual education literature and research is that decent and honorable people who were once reformers have become the conservative "establishment." They have forgotten that the purpose of bilingual education is to help children. It is not an end in and of itself.

BIBLIOGRAPHY

BAKER, K.A., & de KANTER, A.A. (1981). Effectiveness of bilingual education: a review of the literature. Washington, D.C.: U.S. Department of Education.

DANOFF, M.N., ARIAS, B.M., COLES, G.J., & Others (1977). Evaluation of the impact of ESEA Title VI Spanish/English bilingual education program. Palo Alto: American Institutes for Research.

DANOFF, M.N., COLES, G.J., McLAUGHLIN, D.H., & REYNOLDS, D.J. (1978). Evaluation of the impact of ESEA Title VII Spanish/English bilingual education program. Vol III: Year Two Impact Data. Palo Alto: American Institutes for Research.

OKATA, M., et al. (1983). Synthesis of reported evaluation and research evidence on the effectiveness of bilingual education basic projects. Los Alamitos, CA: National Center for Bilingual Research.

ROSSELL, C.H., & ROSS, M.J. (1985). The Social Science Evidence on Bilingual Education. Boston: Boston University Center for Applied Social Science Working Paper 85-7.

ROSSELL, C.H., & ROSS, M.J. (1986). The Social Science Evidence on Bilingual Education. The Journal of Law and Education, 16, Fall 1986.

WILLIG, A.C. (1981-82). The effectiveness of bilingual education: review of a report. NABE Journal, 6, 1-19.

YOUNG, M.D., and Others (1984). LEP student characteristics and school services. Arlington, VA: Development Associates.

ZAPPERT, L.T., & CRUZ, B.R. (1977). Bilingual education: An appraisal of empirical research. Berkeley: Bay Area Bilingual Education League.