DOCUMENT RESUME

ED 207 992 TM 810 646

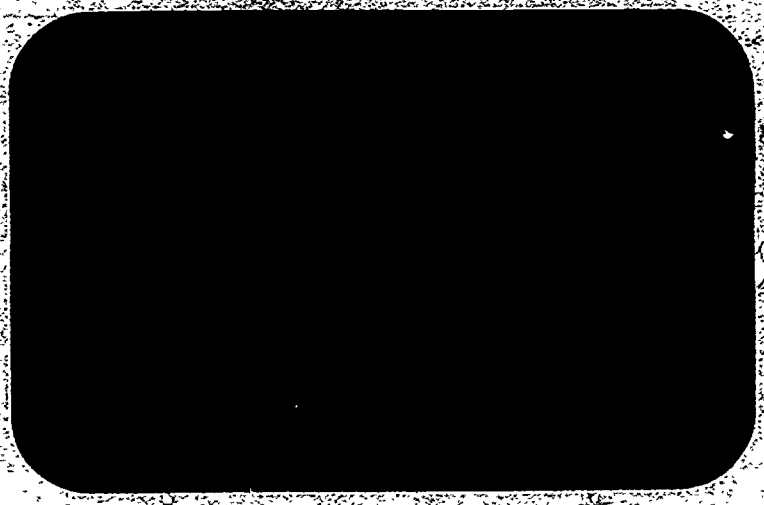
AUTHOR Rosario, Jose; Love, John M.
TITLE Evaluations of Bilingual Programs: Examples of the Reproductive Functions of Evaluative Research. Bilingual Education Paper Series, Vol. 4 No. 7.
INSTITUTION California State Univ., Los Angeles. National Dissemination and Assessment Center.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Feb 81
NOTE 35p.; Revision of a paper presented at a forum on Ethnoperspectives in Bilingual Education Research (Eastern Michigan University, Ypsilanti, MI, April, 1979).

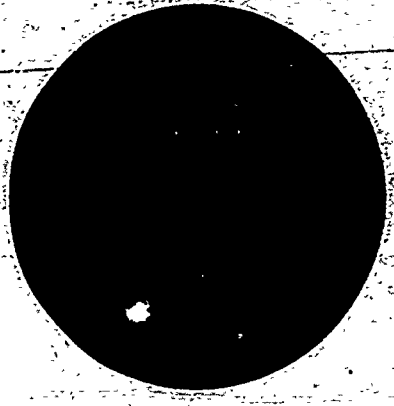
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Bilingual Education; Early Childhood Education; *Evaluation Methods; Policy Formation; Program Design; *Program Evaluation; Spanish Speaking
IDENTIFIERS *Bilingual Programs; *Evaluation Utilization; *Project Head Start

ABSTRACT

Using the American Institutes for Research (AIR) Report (1978) and evaluation of the Head Start Bilingual/Bicultural Curriculum Project by Juarez Associates as contrasting examples of bilingual research, the authors suggest a new direction for policy research in bilingual education oriented towards examining all factors affecting such policy formation within the larger relational nexus. To provide a context for the discussion, the Head Start Strategy for Spanish Speaking Children and the plans for the evaluation of its curriculum development component are described. An outline of the purpose and the design of the evaluation project follows. This involves not only the administration of tests to measure the impact of program objectives, but the collection of data to illustrate the nature and extent of program implementation over time, which can then be used to determine the feasibility of implementing the curriculum models at other sites; and collection of qualitative observation and change data over time for use as outcome data. The reproductive function this evaluation is likely to serve is then exemplified by drawing attention to how evaluative research can be used to legitimate social action programs. Future directions are also discussed. (Author/AEF)

Bilingual Education PAPER SERIES





National Dissemination and Assessment Center
California State University, Los Angeles
Los Angeles, CA. 90032

ISSN 0161-3707

Vol. 4 No. 7
February 1981

EVALUATIONS OF BILINGUAL PROGRAMS: EXAMPLES OF THE REPRODUCTIVE FUNCTIONS OF EVALUATIVE RESEARCH

José Rosario
High/Scope Educational
Research Foundation
Ypsilanti, Michigan

John M. Love
Mediax Associates, Inc.
Westport, Connecticut

The subject of this publication was supported in whole or in part by the Office of Education, United States Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education; no official endorsement by the Office of Education should be inferred.

This publication was printed with funds provided by the Bilingual Education Act, Title VII of the Elementary and Secondary Education Act of 1965, as amended by Public Law 93.380.

EVALUATIONS OF BILINGUAL PROGRAMS: EXAMPLES OF THE REPRODUCTIVE FUNCTIONS OF EVALUATIVE RESEARCH*

José Rosario
John M. Love

Currently, 11 major national evaluation studies of Head Start services are underway. These evaluations

> ...are designed to improve the quality and delivery of local Head Start services in the areas of education, parent involvement, social services, health (including medical, dental, mental health and nutrition), and services for children with special needs--handicapped Spanish-speaking, Indian and migrant. (*Head Start Newsletter*, 1978, p. 6)

One of these 11 major studies is *An Evaluation of the Head Start Bilingual/Bicultural Curriculum Development Project* (Juárez and Associates, 1978). This is a major evaluation, and it should not be taken lightly by those with an interest in bilingual education. Although it has a different purpose and audience, given its scope and design, its impact may very well make the American Institutes for Research (AIR) evaluation of Title VII programs (1978) pale in comparison. As such, it deserves close attention.

Aside from the fact that the Juárez and Associates evaluation (1978) seems more promising and interesting than the AIR evaluation, it may also be worth considering whether we can expect the

---

*This is a revision of a paper presented at a forum on Ethnoperspectives in Bilingual Education Research held at Eastern Michigan University, Ypsilanti, Michigan, April 1979.

Juárez and Associates evaluation to make a difference in shaping early childhood bilingual education policy formation. Although this question seems premature at this point, there is sufficient reason to claim that the findings of the Juárez and Associates evaluation will likely be used as intended. The irony here is that the substantive findings may not really matter. As we see it, the evidence seems too compelling to conclude otherwise. Notwithstanding radical shifts in administrative leadership and government policy, the evaluation findings will probably be used to legitimate federal efforts in the field of early childhood bilingual education. To expect otherwise is to overlook the structural linkage that exists between educational research and the institutions that sponsor it.

The general purpose of this paper is to present this evaluation in an open forum to allow greater discussion of its potential impact on bilingual education. The Juárez and Associates evaluation (1978) offers us the attractive opportunity of specifically looking at the reproductive function of evaluative research, a topic that should be of greater interest to bilingual education researchers.

We argue that evaluative research can and many times does serve a reproductive function and that the Juárez and Associates evaluation (1978) can and probably will serve that function rather well. The argument will proceed as follows: First, in order to shape a context for the discussion, we provide a brief description of the Head Start Strategy for Spanish-Speaking Children and discuss generally the plans for evaluating its cur-

riculum development component. Second, we will discuss the purposes and design of the Juárez and Associates evaluation. Third, we will point to the reproductive function the Juárez and Associates evaluation is likely to serve by drawing attention to how evaluative research can be used to legitimate social action programs. [The Westinghouse/Ohio evaluation (Datta, 1976) will serve as an example.] Finally, we will conclude by trying to carve out some direction for this kind of work.

HEAD START STRATEGY FOR SPANISH-SPEAKING CHILDREN

It is estimated that approximately 19 percent of the total child population currently served by Head Start is Spanish surnamed. This is a sub-population of children with known variation along the dimensions of language use and place of origin, coming from families that are Mexican, Puerto Rican, Cuban, or Latin American in origin. While some speak limited English, many others speak only Spanish.

In 1975, the Office of Child Development--now the Administration for Children, Youth and Families (ACYF)--initiated a major program to address the specific needs of these children. This comprehensive program was named the Head Start Strategy for Spanish-Speaking Children. This strategy consists of carrying out substantial work in four relatively undeveloped areas of bilingual early childhood education: staff training, bilingual/bicultural resource networks, research, and curriculum development (Contracting Corporation of America, 1977).

The current work in basic research, for example, covers such areas as early childhood bilingualism, heart murmurs among Mexican-American children, and parental resistance to diagnostic and remedial heart care. The work in the area of staff training, on the other hand, reduces to the development of four bilingual/bicultural staff training models following the competency-based framework set forth in the Child Development Associate (CDA) program. There is also concurrent development of two additional models specifically designed to include procedures for training bilingual/bicultural CDA trainers. To remedy the lack of bilingual/bicultural resource networks, a prototype network was set up in Denver, Colorado to provide Head Start agencies in Region VIII with in-service training, bilingual/bicultural materials, and technical assistance in the implementation of bilingual/bicultural programs. This prototype network was designed and established to later serve as a replicable model for other regions. The network is now functioning, and the nationwide replication is in process.

Finally, the work in the area of curriculum development focused specifically on the design of bilingual/bicultural curricula for preschool children. In 1976, ACYF contracted with four institutions--Columbia Teachers College, the High/Scope Educational Research Foundation, Inter-Cultural Development Research Associates, and the University of California at Santa Cruz--to take on the task of developing four early childhood bilingual/bicultural curriculum models. The contractors were expected to work closely with local Head Start centers in a four-year develop-

ment process. The work plan for the first year (1976-1977) consisted of model development and preparation for training Head Start staff and parents to carry out implementation of the model, expected to occur during the second year (1977-1978) at the participating Head Start centers in the communities where model development was occurring: New York, New York; Detroit, Michigan; San Antonio, Texas; and Watsonville, California. The third year (1978-1979) called for model replication in Head Start centers at two additional sites. The fourth year (1979-1980) was devoted to the provision of continuing supervision and technical assistance to the Head Start centers at the replication sites.

Thus far, development of the models shows them to be fundamentally similar in principle but dissimilar in educational approach. All the models explicitly acknowledge the importance of attending to child development and the language and cultural needs of Spanish-speaking children as bases for program design, but each model calls for a different programmatic approach to the education of bilingual children. These differences are to be expected, for the pedagogical principle structuring the efforts behind Head Start's bilingual/bicultural curriculum project is the "no single best approach." The director of the Head Start Strategy for Spanish-Speaking Children describes the principle as follows:

> There is no single "best" bilingual/bicultural preschool program. The precise form such a program takes should be a function of the group it serves. The cultural and linguistic differences among Chinese, Native American, Spanish, Filipino and other groups are wide, and there

> are differences even within each group. A preschool program serving Chinese-American children in San Francisco, for example, may not be appropriate for a Chinese-American group in New York. Furthermore, several racial and ethnic groups, with or without a different language or dialect, may be represented in a particular community. Often the reality of a preschool program is that it is bilingual and multicultural. The specific style and content of each program must be tailored to the needs of the community and the groups within it. (Arenas, 1978, p. 3).

Of the four components in the Head Start Strategy for Spanish-Speaking Children, the curriculum development project is the only one being evaluated; and plans for evaluating the products of this "no single best approach" principle were set in motion in 1977. ACYF contracted with Juárez and Associates (a consulting firm based in Los Angeles, California) to conduct an evaluation of the Head Start Bilingual/Bicultural Curriculum Development Project. The evaluation was to unfold in four phases over a three-and-a-half year period. The first phase (September 1977-October 1978) was devoted to selection of the measurement battery and the sites expected to participate in the evaluation (replication sites of model developers). In addition, the evaluators were supposed to collect initial pretest data from teachers. During the second phase (October 1978-September 1979), the evaluators were expected to pilot test the measurement battery and randomize subjects into experimental and control groups. Collection and analyses of data were to be performed during the third phase (1979-1980). Finally, the fourth phase (October 1, 1980 to March 31, 1981) was to be devoted to writing a final report and

shaping the knowledge distribution system used to inform program decision-making at the local Head Start level: a set of pamphlets that describe each model in detail and discuss, among other things, requirements for its implementation.

GOALS OF THE JUÁREZ AND ASSOCIATES EVALUATION

The guiding purpose of the Juárez and Associates evaluation (1978) was to supply ACYF with information on the effectiveness of four early childhood bilingual/bicultural models for Spanish-speaking children. Once the information was supplied, ACYF planned to use it in deciding on the feasibility of wider distribution of the models to other Head Start centers across the nation. This basic information need on the part of ACYF required that the evaluation focus on the collection of three types of data.

First, the evaluation must find out whether it is feasible to successfully implement the curriculum models in more than one setting. To address this question, the evaluators collected information bearing on two general factors related to program implementation: the *process* required for installing the program, and the program *procedures* required of the model for activating its valued message systems.[1] Among the models' procedural features on which the evaluators expect to focus are staff, community, resource, and student characteristics required by the model for

[1] The concept of message systems is a Bernsteinian notion that is used to refer to curriculum, pedagogy, and evaluation. For a detailed discussion, see Bernstein, 1975, pp. 85-115.

it to function as designed.

Second, the evaluation must find out whether the model objectives are being met by measuring impact on children, teachers, and parents. The contractual agreement is explicit as to how this kind of measurement must be carried out. ACYF has required the evaluation to employ an experimental pre- and posttest design in measuring model impact. At each replication site, a total of 90 four-year-old children were to be randomly assigned to treatment and control groups. Following this random assignment, the children would be pre- and posttested on a selected number of competency measures: (1) Spanish language comprehension, (2) Spanish language production, (3) English language comprehension, (4) English language production, (5) concept development, and (6) socio-emotional development.

To measure model impact on teachers, pre/post interviews were planned, designed to tap background characteristics and attitudes toward, knowledge of, and competency in early childhood bilingual/bicultural education. The same procedure was used to measure model impact on parents. The plans included pre/post administration of a questionnaire designed to tap background characteristics and attitudes toward, and knowledge of, general education, bilingual/bicultural education, and vocational careers for their children.

Lastly, the evaluation must find out whether Head Start staff, parents, and other community members have received the curriculum models favorably. This information need was met by gauging staff and parental satisfaction with the model. Parents, for

example, were expected to comment on the degree to which the curriculum models were compatible with their views of how their children should be educated. Teachers, on the other hand, would be asked to report on the quality, clarity, suitability, and perceived effectiveness of the model.

A goal of all this research activity is the production of pamphlet summaries describing the evaluation findings bearing on each model. ACYF will then use these pamphlets to inform local Head Start centers of the options available to them when deciding on how best to meet the special needs of Spanish-speaking children. The pamphlets would describe each model, discuss what each would require for implementation, identify what each would offer in the form of programmatic strengths, and provide data on how effective each model was found to be with children, parents, and teachers. It is not the intent of the pamphlets to provide a comparative analysis of four models, since such an analysis of the models *à la* "Follow Through" was not among the goals of the evaluation. The model comparisons were being left to the consumers of the models.

DESIGN OF THE JUÁREZ AND ASSOCIATES EVALUATION

The design of the Juárez and Associates evaluation (1978) reduces to what they call a "holistic or multi-method data collection strategy." It has also been referred to as a "contextual evaluation study design." For evaluators, these terms are interchangeable; and they all imply an expansion of the experimental approach so as to include ratings, event sampling, and ethnographic techniques. The inclusion of these other approaches

into the evaluation design specified by the evaluation contract, the argument goes, would allow for drawing relationships between program processes (input) and outcome measures (output). This argument stems from four key issues related to evaluative research. All four issues relate to the need of having clear and exact knowledge of: (1) program treatment, (2) program participants and settings, (3) program contexts, and (4) control group activities. In isolating these four issues, Juárez and Associates (1978) draw on a growing consensus among researchers that the lack of these data is likely to make research findings (particularly psychometric test results) uninterpretable. This is the kind of reasoning that has helped shape the design of the so-called contextual evaluation study of Juárez and Associates.

There are three basic objectives to this type of study design. The first is to measure the nature and extent of program implementation over time in order to demonstrate the feasibility of implementing the curriculum models at varying sites. Specific data on setting and material resources, individuals, schedule and program organization, attitudes and actual program activities were to be included in analyses employing a *fidelity* perspective in the measurement of model implementation.[2]

The second objective of the design is to secure the kind of observational data of use in interpreting test outcomes. In this

---

[2]Although the intent here is not to provide a critique of the evaluation design, there is a compulsion, however, to note that the *fidelity* perspective as used traditionally in measuring implementation is problematic in securing accurate assessments of institutional change. See for example, Berman and McLaughlin (1974).

case, the qualitative data becomes an explanatory adjunct to quantitative results. The more interesting feature here is the varying kind of qualitative data that Juárez and Associates (1978) targeted for collection:

> Observations of classroom activities and situations, such as specific lessons, and language use in various situations, will be taken into consideration in the interpretation of test data. These observations will also include information on the measurement process or how the children in both the treatment and the control groups were tested, and on the experimental arrangement, including information on the control group experience and how well randomization worked. All of the above activities may have a strong influence on test results. (I-p.10)

The third objective is to secure qualitative change data over time. The intent here is not to use qualitative data to help explain test results but to use qualitative data itself as outcome data. As Juárez and Associates view it (1978), the use of qualitative data as outcome data is a complementary and equally valid way of providing information bearing on model impact. The strategy has the added and very attractive feature of being able to tap unanticipated outcomes. This is an important point, for it illustrates the sensitivity of the evaluation toward the very real and commonly acknowledged possibility of side effects due to explicit as well as *implicit* features of program designs (Monaghan, 1976). Although important, the measurement of program side effects is an item routinely omitted by evaluation designs. In sum, among the foci of the qualitative data collection strategy were specific classroom behaviors and control group activities.

To operationalize this multi-method design, the evaluators proposed the use of a "participant researcher" (PR). This PR was seen as very instrumental to the Juárez and Associates (1978) scheme. One reason for suggesting the presence of a PR on site was the need to eliminate the "outsider" image of evaluators. In a sense, PR might very well stand for public relations in this case. By becoming an insider, the PR was expected to get a deeper view of program evaluations. While onsite, the PR was expected to supervise psychometric testing, conduct naturalistic observations, administer interviews to teachers, and update treatment and control group information through periodic phone calls to parents. As proposed, the use of the PR is an interesting and intriguing innovation in evaluation studies that reflects the strong influences that ethnography is having on evaluative research generally and on the Juárez evaluation in particular.

To summarize, the Juárez and Associates (1978) evaluation design contains many features critical to successful evaluations--features lacking in other federally sponsored evaluations of bilingual education. There have been four well known federally sponsored studies: (1) *A Process Evaluation of the Bilingual Education Program, Title VII, ESEA* (Development Associates, 1973); (2) *Bilingual Education: An Unmet Need* (Comptroller General of the United States, 1976); (3) *The Condition of Bilingual Education in the Nation* (United States Commissioner of Education, 1976); and (4) *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program* (Danoff, 1978). Differences in purpose, range, and focus make it difficult to compare these evalua-

tions and the Juárez and Associates (1978) study. Nevertheless, a brief summary of one of the most widely cited of these studies (the AIR Report, 1978) will help to place the Juárez study in context. The Juárez effort represents an impact study of bilingual education with many of the features hoped for in well-designed evaluations. We will examine how it differs from the Title VII evaluation and then explore how these differences increase the opportunity for the current study to serve a true reproductive function.

THE AIR REPORT

The American Institutes for Research (AIR) conducted a major evaluation of Title VII programs. Interim results were released in April, 1977; and the Executive Summary of the final report was released by the Office of Education in March, 1978. These reports were soon widely cited as evidence that bilingual education was not working. When the next year's education bill was debated, Congressman Ashbrook of Ohio proposed an amendment to abolish bilingual education, saying that "the program is actually preventing children from learning English." The evaluation findings were used to support some lawmakers' views of public policy; but, for the most part, Congress has been supportive of bilingual education, increasing the budget annually.

In the fall and spring of 1975-1976, thousands of students in grades 2-6 were tested at 38 different sites across the country (in a total of 150 schools). The Title VII projects included in the study were in their fourth or fifth year of funding and pre-

sumed to be mature bilingual programs. The comparison group was selected by personnel who each nominated one non-Title VII classroom with students comparable to each Title VII classroom. The match was in terms of ethnicity, socio-economic status, and grade level. AIR reported that the non-Title VII students were basically comparable to the Title VII students and that standard statistical procedures were used to adjust for differences between groups. Standardized achievement tests were used to measure English and Spanish Oral Comprehension and Reading, and Mathematics. A student questionnaire obtained student background information and attitudes toward school-related activities. A sub-sample of classrooms were observed and teachers and directors interviewed to obtain information on the educational experiences of the students (Danoff, 1978).

Three different methods of analysis were used to assess Title VII's impact. Overall, "...the Title VII Program did not appear to be having a consistent significant impact on student achievement in these two subjects [English Language Arts and Mathematics]" (AIR Report, 1978, p. 17). The comparisons with non-Title VII children showed that Title VII students were performing worse in English than non-Title VII students. In Math, the two groups were performing at about the same level. An increase in Spanish Reading scores was found during the year for Title VII students, but these gains could not be compared with the non-Title VII students because the number who could complete the test was too small.

The study also examined the educational goals of Title VII programs. In interviews with directors of Title VII programs, 86 percent reported that children remained in the bilingual project even after they were able to function in English. According to the AIR Report, "[t]hese findings reflect Title VII project activities which run counter to the 'transition' approach strongly implied by the ESEA Title VII legislation" (1978, p. 12).

One of the major difficulties of most large-scale educational evaluations is locating and maintaining an appropriate comparison group. We have the word of the AIR researchers that "classrooms" were matched in terms of SES, ethnicity, and grade level. Yet the students in Title VII programs did worse on tests of English than their matched comparisons. Did AIR match the groups on initial language facility or language dominance? The report doesn't say, but it was reported that the increases in Spanish Reading scores of the Title VII students could not be compared with the non-Title VII student's scores because too few of the latter could even take the Spanish reading test.

Groups were judged for comparability in 1975-1976, not when the students began their Title VII program. The study also grouped all Title VII programs together with no attempt to identify and separately analyze quality bilingual programs.

Perhaps one of the major weaknesses of the study is an ambiguity in its purpose. The United States Office of Education sought to discover the impact of Title VII programs (primarily on the children) when a prior question might have been, "What educational programs are being implemented under the Title VII legis-

lation?"

The politics of bilingual education in the United States is not unrelated to the politics of "basic skills," and this climate affects reaction to the study. In the current debate regarding a return to the "basics," one hears reference to the traditional American values inherent in the mechanics of reading, writing, and arithmetic. In debates on bilingual education, there is the parallel reference to the virtues of the traditional "melting pot" process in America. It may not be the achievement findings that cause concern among educators but the language maintenance approach reflected in the responses of the Title VII program directors that concerns opponents of bilingual education.

THE REPRODUCTIVE FUNCTION OF THE JUAREZ EVALUATION

There is no such thing as neutral research. Whether of the order of scientific interests Habermas (1971) describes, or ideology, there are specific interests underlying all forms of knowledge production. Apple (1977) describes the problem well:

> ...what is the prevailing function of research? Isn't it merely a process of data production that tries to help us solve our day-to-day problems? I am afraid that our answer here must be no, for an affirmative answer neglects one important yet too often forgotten social fact: research creates information for use *by* somebody. It is sponsored by and affiliated with organizations that have a stake, though often not a conscious one, in the continued maintenance of their and other more powerful institutions' basic modes of operation. Thus, we must ask the political question, "Why is this data produced?" One should not conceive of the products of educational research apart from the institutional functions they perform. Too often they perform a rather

> interesting role; they act to prevent committed educators from focusing on the contradictions within our institutions by forcing attention on official definitions of problems. (p. 118)

Evaluative research is no exception to this. Indeed, nowhere is the reproductive function of research more apparent than in evaluative studies, for these essentially seek a measurement of the reproductive value of that which they study. Now that social systems generally look to scientific rationality for legitimacy (Habermas, 1970), the reproductive function of evaluative research becomes increasingly important. The national evaluation of Project Head Start is illustrative of just how important a role evaluative research can come to play in the social reproduction of a major federal program.

The controversy created by the Westinghouse/Ohio Evaluation of Project Head Start (Datta, 1976) is well known. The political and methodological issues that it raised were many, and the debate was heated and long. But now that the controversial dust has settled and a number of significant events in the history of the evaluation have been reconstructed, we can see clearly that the Westinghouse/Ohio evaluation seems to have served its function well: it helped to preserve Project Head Start and deliver to its leadership what it had desired all along.

The most important feature of the Westinghouse/Ohio evaluation (Datta, 1976) was not the set of findings that were generated but the *scientific legitimacy* that it produced for the internal and external expansion of Project Head Start. That Head Start

summer programs were generally less effective than full-year programs and that the impact on Head Start children tended not to persist into the primary grades was of no surprise to the Head Start Project. As Datta (1976) points out:

> Evidence suggesting that summer programs were less developmentally effective than full-year programs and that the academic gains of Head Start were not sustained had been available almost since the beginning of Head Start.... The greater benefits of full-year versus summer programs were evident also in the 1965-to-1969 data...most advisors...cautioned that brief programs would have limited effects on language, cognition, or other aspects of intellectual development. (pp. 144-146)

The fact that this knowledge was available to Project Head Start prior to its evaluation helps to explain why the Head Start Research Council argued strongly against the design of the Westinghouse/Ohio study. Although it was unable to alter the basic design of the impact study, the Head Start Research Council at least managed to influence the study to some degree. Measures of parental satisfaction and child affect and motivation became part of the measurement battery. To complement the study, an advisory board of experts was also established.

The reaction of the Head Start Council indicates that the evaluation of the project may have been feared. In light of the knowledge then available to Head Start, this fear seemed justified. Nevertheless, there remained a need to legitimate expansion of the Head Start project. Such legitimation came from findings and recommendations supposedly based on scientific rationality. As sources of this rationality, the roles played by the Westinghouse Learning Corporation and Ohio University can-

not be under-estimated. A more accurate understanding of the Westinghouse/Ohio evaluation requires relational analysis: that the report not be examined apart from its producers--two institutions interested and committed to preserving and expanding research.

That the evaluation served a reproductive function is suggested by the Office of Child Development leadership as it reconstructs the history of the Westinghouse/Ohio evaluation:

> We used Westinghouse and other studies constructively to reaffirm the Head Start focus on the whole child; the concerns about the summer projects were reasserted and the short-term innoculation notion was shown to be fallacious; the notion of a single magic year was also shown to be limited; and the folly of selecting as criteria stable measures of development was emphasized.
>
> Westinghouse is just the only book on the shelf about Head Start and so people point to it. But the issues debated about Head Start and early childhood education are issues we'd have been debating anyhow.
>
> Westinghouse helped justify what we wanted as early as 1966: continuity, earlier involvement, greater concern about parent involvement. And we used Westinghouse to legitimize these. (Datta, 1976, p. 151)

It is true that the Head Start budget stabilized after the report. It is also true that Elliot Richardson, Secretary of the Department of Health, Education, and Welfare under Richard Nixon, may have received convincing pressure from parents to preserve Head Start. But there are two factors we need to keep in mind. First, the demonstration of Head Start parents in Washington buttressed one of the few positive findings of the Westinghouse/Ohio evaluation (Datta, 1976): parental satisfaction, added as a measure to the study at the insistence of the Head

Start Research Council. In a sense, the report may have functioned to legitimate the action of the parents as well. Second, policies in Washington were beginning to shift as a result of new administrative leadership. The untimeliness of the report almost interfered with its reproductive function. The new leadership at the Department of Health, Education, and Welfare found itself holding the results of a study it had neither sponsored nor been prepared to use. The evaluation had losts its political legitimacy--a case of scientific rationality without political backing. The political legitimacy of the evaluation had to be reconstituted in the form of both federal and grass roots support.

In the case of the Juárez and Associates evaluation (1978), there is a key factor strongly favoring its reproductive function--the *structural linkage* that exists between the Administration for Children, Youth and Families (ACYF) and the evaluation. On the one hand, ACYF wants to help Head Start programs desiring to implement bilingual curricula to "be able to select the model that best suits their local needs and is compatible with their staff's teaching style" (Arenas, 1978-1979, p. 18). This implies a very different use of information than making budget decisions about a national bilingual education program like Title VII. On the other hand, a consulting firm consented to address the "research" problem that ACYF had defined and to design and carry out an evaluation to resolve it. The result was an evaluation specifically designed to generate findings that local Head Start programs could survey to judge not only how well a given curriculum model worked but under what "circumstances and needs" each worked or failed to work.

Although the AIR Report (1978) obtained some program descriptive data, this information was not sufficiently precise to permit an assessment of existing circumstances and under which ones particular effects would be found. Thus, we are left at the end of the AIR Report with basically one global conclusion: Title VII children do no better, and perhaps worse, than their non-Title VII counterparts. This is the kind of conclusion that tends, at least politically, to weaken the reproductive power of an evaluation. When it happens, the loss in reproductive power then has to be restored through recommendations of some kind. Because of the different purpose and the more "open" design of the Juárez and Associates evaluation (1978), the reproductive function of that evaluation may not need the doctoring usually found in evaluations in the form of "positive" recommendations suggesting that the program be saved.

The structural linkage between ACYF and the Juárez and Associates evaluation (1978) makes the study more focused and more limited, focusing on a particular set of curriculum projects that were funded for a limited time period and limited in size (four sites and only hundreds of thousands of dollars spread over several years). In no way does it approach the multi-million dollar Title VII program. The Juárez and Associates findings can be used by a single agency to shape its own strategy toward bilingual children and families and does not have to serve as "the" evaluation of "United States bilingual education."

Furthermore, ACYF has a history of supporting program development and research efforts, raising the probability that findings from the Juárez and Associates study will be received positively

and acted on by the agency. In 1972, for example, ACYF (then the Office of Child Development) began Home Start, a home-based variant of Head Start and simultaneously funded an evaluation designed to collect both process and outcome data for judging the implementation and impact of the program. Information from the evaluation was used in modifying guidelines for the program; and when Home Start ended and more than 300 Head Start programs adopted a home-based option (usually to supplement their ongoing center-based program), ACYF instituted six regional training centers to provide training and technical assistance to Head Start programs making this change (Love, 1978). If the Juárez and Associates evaluation (1978) provides useful information about the four bilingual curriculum projects, there is every indication that ACYF will attempt to base future programmatic decisions on that information.

It seems to us that mandates to evaluate institutional policies and practices presents a troublesome dilemma to policymakers, especially those with an interest in preserving certain policies and practices considered important and worthwhile. The policymaker has to demonstrate the worth of a given policy or program to justify its preservation. How does the policymaker do this without jeopardizing the policy or program? One option is for the policymaker to realistically and clearly define what he/she wants the evaluation to do and then support a well thought-out evaluation design capable of delivering the kind of information usable in maintaining or improving that policy or program. That is the posture reflected in the Juárez and Associates evaluation

(1978), and we submit that it is as reasonable as any other.

TOWARD A NEW DIRECTION IN BILINGUAL/BICULTURAL EDUCATION POLICY RESEARCH

There is a dire need in bilingual/bicultural education policy research for engaging in *relational analysis*. This perspective requires that all social activity be construed as structurally related to other forms of social activity and institutions. Social actions are defined in terms of their relations to other forms of social actions and underlying principles of social organization and control. Apple (1979) describes this research approach as follows:

> What I am asking for is what might best be called "relational analyses." It involves seeing social activity--with education as a particular form of that activity--as tied to the larger arrangement of institutions which apportion resources so that particular groups and classes have historically been helped while others have been less adequately treated. In essence, social action, cultural and educational events and artifacts (what Bourdieu would call cultural capital) are "defined" not by their obvious qualities that we can immediately see. Instead of this rather positivistic approach, things are given meaning rationally, by their complex ties and connections to how a society is organized and controlled. The relations themselves are the defining characteristics. Thus, to understand, say, the notions of science and the individual, as we employ them in education especially, we need to see them as primarily ideological and economic categories that are essential to both the production of agents to fill existing economic roles and the reproduction of dispositions and meanings in these agents that will "cause" them to accept these alienating roles without too much questioning. They become aspects of hegemony. (p. 10)

The absence of this view in bilingual education research is a bit ironic. As reformers, bilingual educators have done their share to accentuate and eliminate inequities in educational opportunity. As researchers, however, they have not given equal attention to discovering the structural reasons for these inequities. The tendency has been to rely more on developing languages of incompatibilities and bicognitivism to justify institutional amelioration rather than restructuration: that is, with legitimating the improvement of practice as opposed to creating different structural arrangements that are more responsive. The problem here is that these language systems function more as slogans and tend to fall short of what is needed for understanding the structural relations underlying social and cultural reproduction.

What is needed is a different metaphor or language system that allows us to describe the structural relations between institutions and individuals so that we can see how individuals and institutions function for and against each other. The kind of language system we have in mind here is central to the sociology of school knowledge being developed by Willis (1977), Eggleston (1977), Young (1971), Bernstein (1975), Bourdieu and Passeron (1977), Green and Sharp (1975), and Apple (1979).

In the work of these theorists we find a metaphor virtually absent from the literature on bilingual education. This metaphor links educational practice, including research, to other social processes and institutions in the society. Education, for example, is seen as essentially a mechanism of cultural preservation and distribution that helps to create and recreate forms of conscious-

ness. Education is thus the means by which a society reproduces itself. The key here, though, is that this reproduction is subtle, remaining largely unconscious on the part of educators, rather than as the result of an elite group of managers who sit around tables plotting ways to "do in" their workers in the workplace and school. Educational activity and interaction are so thoroughly tied to the patterns of the wider society that they are outside the conscious awareness of educators. According to Apple (1979), the Gramscian construct of *hegemony* best describes the relationship between the assumptions and behaviors of educators on the one hand and the structure and patterns of the wider society on the other.

Basically, *hegemony* implies that the fundamental economic and political divisions in a society

> ...are held together by tacit ideological assumptions, or rules, which are not usually conscious. . . These rules serve to organize and legitimate the activity of the many individuals whose interaction makes up a social order" (Apple, 1979, p. 86)

The assumptions necessary for maintaining the existing social order so thoroughly saturate the activities of those within the society that they actually define what is common sense; individuals behave in a manner that reinforces the social order, not because they are forced but because they cannot conceive of any other way to act. In the realm of education, Apple (1979) sees hegemony manifested in the deployment by educators of *intellectual traditions* for both legitimating and transmitting existing practices. He calls this phenomenon "ideological saturation." It can be illustrated with analyses of how the true role of conflict in

social change is glossed over in science curricula and by exploring the language of diagnosis and labeling used in elementary schools. In both cases, Apple shows how these "selective traditions" are used to control and ultimately victimize whole classes of children. Again, Apple is careful to warn against construing his argument to imply conscious conspiracies among educators.

> This is not to imply that all school people are racist (though some may in fact be) or that they are part of a conscious conspiracy to "keep the lower classes in their place." In fact, many of the arguments for "community" and about curriculum put forth by some of the early educators, curriculum workers, and intellectuals...were based on the best liberal intentions of "helping people." Rather the argument being presented here is that "naturally" generated out of many...educators' common-sense assumptions and practices about teaching and learning, normal and abnormal behavior, important and unimportant knowledge, and so forth are conditions and forms of interaction that have latent functions. And these latent functions include some things that many educators are not usually aware of. (Apple, 1979, pp. 64-65)

The metaphor of education as cultural reproduction can help us better understand how evaluative research is constrained by and contributes to the maintenance of existing sociocultural patterns. Consider the debate over the Title VII findings, for instance. Throughout this debate, we have somehow overlooked the critical issue that has been raised: whether the federal government should play a mediating role in the reproduction of ethnic languages and cultures. We can't respond to that issue by pointing to a faulty evaluation design. The issue is politically charged, and research design has very little to do with it. The issue relates more to a structural relation that federal

support for bilingual education seems to be endangering--a political concern that the Title VII evaluation helped to exacerbate in showing how bilingual programs were being used for language and cultural maintenance.

To alleviate the concern, federal policymakers have started a search for entry-exit criteria applicable to Title VII students. This search is certainly a good indicator that the Title VII evaluation has served its reproductive function rather well in certain areas. The findings have so far led to tighter monitoring and control that are, in a sense, improvements and therefore preservation of the program. It makes sense, therefore, to pay less attention to questions of faulty design and more attention to the structural reasons underlying the Title VII evaluation; when looked at closely, the problem with the AIR evaluation has more to do with the purpose and type of questions raised than with faulty design.

To summarize, using the AIR Report (1978) and the Juárez and Associates (1978) evaluations as contrasting examples of bilingual research, we have suggested a different direction for policy research in bilingual education. We suggest an orientation that looks at all the factors (program evaluations being one example) affecting bilingual education policy formation within the larger relational nexus of which they are a part. If these factors are to be accurately interpreted, we cannot afford to settle for anything less.

REFERENCES

American Institutes for Research. "Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program," *Bilingual Education Paper Series*. Vol. 2, No. 1. Los Angeles, California: National Dissemination and Assessment Center, August, 1978.

Apple, Michael W. *Ideology and Curriculum*. London: Routledge & Kegan Paul, 1979.

_____. "Politics and Research," *The Curriculum and Cultural Reproduction*, ed. M. MacDonald. London: Open University Press, 1977.

Arenas, Soledad. "ACYF's Efforts to Address the Needs of Spanish-Speaking Children," *Bilingual/Multicultural Early Childhood Education: Proceedings of Head Start Regional Conferences, 1978-79*, eds. Kenneth A. Martinez, and Soledad Arenas. Washington, D. C.: Administration for Children, Youth and Families, 1978-1979.

_____. "Bilingual/Bicultural Programs for Preschool Children," *Children Today*, VII, No. 4 (July-August, 1978), 2-6.

Berman, Paul, and Milbrey W. McLaughlin. *Federal Programs Supporting Educational Change: A Model of Educational Change*. Vol. 1. Prepared for the United States Office of Education, Department of Health, Education, and Welfare. R-1589/1-HEW. Santa Monica, California: Rand Corporation, 1974.

Bernstein, Basil. *Class, Codes and Control*. Vol. III. London: Routledge & Kegan Paul, 1975.

Bourdieu, Pierre, and Jean-Claude Passeron. *Reproduction in Education, Society and Culture*, trans. Richard Nice. London: Sage Publications, 1977.

Comptroller General of the United States. General Accounting Office Report to Congress. *Bilingual Education: An Unmet Need*. No. MWD-76-25. Washington, D. C.: Government Printing Office, 1976.

Contracting Corporation of America. *Bilingual/Bicultural Preschool Projects*. San Antonio, Texas: Contracting Corporation of America, 1977.

Danoff, Malcolm N. *Evaluation of the Impact of ESEA Title VII Spanish/English Bilingual Education Program: Overview of Study and Findings.* Palo Alto, California: American Institutes for Research, 1978.

Datta, Lois-Ellin. "The Impact of the Westinghouse/Ohio Evaluation on the Development of Project Head Start: An Examination of the Immediate and Long-Term Effects and How They Came About," *The Evaluation of Social Programs*, ed. Clark C. Abt. Beverly Hills, California: Sage Publications, 1976, pp. 129-181.

Development Associates. *A Process Evaluation of the Bilingual Education Program, Title VII, ESEA.* Washington, D. C.: Development Associates, 1973.

Eggleston, John. *The Sociology of the School Curriculum.* London: Routledge & Kegan Paul, 1977.

"Evaluation Studies of Head Start Services," *Head Start Newsletter*, (November-December, 1978), 6-7.

Green, Anthony, and Rachel Sharp. *Education and Social Control.* London: Routledge & Kegan Paul, 1975.

Habermas, Jürgen. *Knowledge and Human Interests*, trans. Jeremy J. Shapiro. Boston, Massachusetts: Beacon Press, 1971.

_____. *Toward a Rational Society: Student Protest, Science, and Politics*, trans. Jeremy J. Shapiro. Boston, Massachusetts: Beacon Press, 1970.

Juárez and Associates. *An Evaluation of the Head Start Bilingual/Bicultural Curriculum Development Project.* Los Angeles, California: Juárez and Associates, 1978.

Love, John M. "The Impact of the Evaluation of the National Home Start Demonstration Program," *The Impact of Evaluation: Lessons Drawn From the Evaluations of Five Early Childhood Programs*, eds. A. C. Granville *et al.* Symposium presented at the annual meeting of the American Educational Research Association, Toronto, Canada, 1978.

Monaghan, A. C. *A Consideration of the Feasibility of a National Study of the Impact of Follow Through on the Public School System.* Cambridge, Massachusetts: The Huron Institute, 1976.

United States Commissioner of Education. *The Condition of Bilingual Education in the Nation.* Washington, D. C.: United States Department of Health, Education, and Welfare, 1976.

Willis, P. *Learning to Labour: How Working Class Kids Get Working Class Jobs.* Lexington, Massachusetts: D. C. Health, 1977.

Young, Michael. *Knowledge and Control.* London: Collier-Macmillan, 1971.

José Rosario received his training at the University of Wisconsin at Madison where he earned a BA in sociology and graduate degrees in curriculum and instruction. He is now a research associate in the Research Department at the High/Scope Educational Research Foundation in Ypsilanti, Michigan, where he is co-principal investigator in an NIE-funded study of stability and change in a public middle school.

Before joining High/Scope, Dr. Rosario spent three years teaching at the university level: one year at Inter-American University of Puerto Rico; and two years at the State University of New York, Geneseo.

Among his publications are contributions to Schooling and the Rights of Children (ed. V. Haubrich and M. Apple), Curriculum Criticism (ed. G. Willis), the Journal Curriculum Theorizing, the Journal of Aesthetic Education, and Curriculum Inquiry.

John M. Love is Vice President for Research & Development at Mediax Associates, an educational research and consulting firm in Westport, Connecticut. He received his PhD in child development from the University of Iowa. Before joining Mediax, he was Director of Research at the High/Scope Educational Research Foundation in Ypsilanti, Michigan.

Dr. Love's main interest is the research measurement, and evaluation of early childhood programs. He has directed several national evaluations of Head Start-related programs, including the evaluation of the Home Start Demonstration Program and the Project Developmental Continuity Evaluation, both of which included important bilingual components.