

# **Project Follow Through:**

**A Case Study of Contingencies**

**Influencing Instructional Practices**

**of the Educational Establishment**

Cathy L. Watkins

**Cambridge Center for Behavioral Studies**  
Cambridge, MA

Copyright © 1997 by the Cambridge Center for Behavioral Studies. All rights reserved.

Permission to reproduce in whole or in part may be granted upon application to the  
Cambridge Center for Behavioral Studies.

Copies of this monograph may be ordered from  
Cambridge Center for Behavioral Studies  
Publications Office  
336 Baker Avenue, Concord, MA 01742.  
Tel: 978-369-CCBS  
Fax: 978-369-8584  
Email: [center@behavior.org](mailto:center@behavior.org)  
Web site: <http://www.behavior.org>

*The Cambridge Center for Behavioral Studies wishes to thank Dr. Beatrice H. Barrett for  
her generous support for the production and distribution of this publication.*

AUTHOR'S NOTE: Please address all correspondence concerning this monograph to:

Cathy L. Watkins, Ph. D.  
Department of Advanced Studies in Education  
California State University, Stanislaus  
801 West Monte Vista Avenue  
Turlock, CA 95382  
Email: [cwatkins@toto.csustan.edu](mailto:cwatkins@toto.csustan.edu)

Printed in the United States of America.

ISBN 1-881317-04-8

Library of Congress Catalog Card Number: 97-78162

# TABLE OF CONTENTS

<b>PREFACE</b> .....	<b>iii</b>
<b>ABSTRACT</b> .....	<b>vii</b>
<b>PART I: HISTORY OF PROJECT FOLLOW THROUGH</b> .....	<b>1</b>
INTRODUCTION .....	1
ORIGINS .....	2
PLANNING.....	6
THE DECISION TO EXPERIMENT .....	7
<i>Budget Cuts</i> .....	7
<i>White House Task Force</i> .....	8
<b>PART II: EXPERIMENTAL DESIGN</b> .....	<b>13</b>
PLANNED VARIATION .....	13
THE SPONSORED MODEL APPROACH.....	15
<i>Selection of Sponsors</i> .....	16
<i>The Follow Through Models</i> .....	18
<i>Selection of Sites</i> .....	19
<i>Pairing of Sites and Sponsors</i> .....	20
<i>Comparison Groups</i> .....	22
SUMMARY.....	22
<b>PART III: EVALUATION</b> .....	<b>25</b>
THE NATIONAL EVALUATION.....	25
<i>A Select History</i> .....	25
<i>Description of Models by Type</i> .....	29
<i>Outcome Measures</i> .....	30
<i>Analytic Strategy</i> .....	32
<i>Summary of Effects</i> .....	32
ALTERNATIVE ANALYSES.....	37
<i>The House Critique</i> .....	37
<i>Bereiter and Kurland</i> .....	41
<i>Kennedy</i> .....	43
OVERALL RESULTS .....	44
<b>PART IV: USING THE RESULTS</b> .....	<b>47</b>
DISSEMINATION .....	47
FUNDING DECISIONS.....	48
<b>PART V: THE EDUCATIONAL ESTABLISHMENT</b> .....	<b>53</b>
POLICY MAKERS.....	53
<i>Role of the Evaluation</i> .....	53
<i>Influence of Stakeholders</i> .....	53
<i>Expectations</i> .....	56
COLLEGES OF EDUCATION .....	59
<i>Conflicts With Existing Philosophies</i> .....	59
<i>Inadequate Skills</i> .....	62
<i>Lack of Effective Contingencies</i> .....	63

PUBLISHERS .....	64
<i>Economic Contingencies</i> .....	64
<i>Instructional Designers</i> .....	66
SCHOOL DISTRICTS .....	66
<i>Financial Contingencies</i> .....	66
<i>Local Norms</i> .....	68
<i>Organization</i> .....	68
<i>Monitoring and Feedback</i> .....	70
TEACHERS .....	71
<i>Teacher Behavior and Achievement</i> .....	71
<i>Inadequate Training and Supervision</i> .....	73
<i>Conflicts With Existing Views of Teaching</i> .....	74
<i>Faulty Contingencies</i> .....	75
THE PUBLIC .....	76
<i>Uninformed About Research</i> .....	76
<i>Misinformed by Educators</i> .....	78
<i>Cultural Relativism</i> .....	79
<i>Failure to Recognize the Scope of the Problem</i> .....	80
<b>PART VI: SUMMARY AND CONCLUSIONS .....</b>	<b>83</b>
PROBLEMS WITH PROJECT FOLLOW THROUGH .....	83
<i>As an Experiment</i> .....	83
<i>Evaluation</i> .....	83
<i>Summary</i> .....	84
IMPACT ON EDUCATION .....	85
<i>What Was Done</i> .....	85
<i>Factors That Influence the Educational Establishment</i> .....	86
<i>Solving The Instructional Problem</i> .....	88
<i>Colleges of Education</i> .....	92
<i>How to Proceed</i> .....	93
CONCLUSION .....	96
<b>REFERENCES.....</b>	<b>99</b>

## PREFACE

When I was an undergraduate student at California State University, Stanislaus, I enrolled in a course taught by Jane Howard. One of the assigned readings was Wes Becker's 1978 Education and Urban Society article, "The National Evaluation of Follow Through: Behavior-Theory Based Programs Come Out on Top." Thus began my interest in Project Follow Through. After completing my Master's degree, I went to the University of Florida to pursue my doctorate. Before long, the time came to choose a topic for my qualifying paper. I am grateful to Jim Johnston for suggesting that I write about Project Follow Through.

Although much has been written about Project Follow Through, little has been widely disseminated. One major reason for writing this paper was to synthesize the vast amount and variety of information and to make it available to an audience that might not otherwise come into contact with it. Another major reason was to try to understand the way that the educational system behaves with respect to instructional methods.

For over a year I spent much of my time in the basement of the library locating and reading documents about Follow Through. One of the most remarkable discoveries I made was Follow Through: Decisionmaking in a Large-Scale Social Experiment, an unpublished thesis by Richard Elmore. Although I have never met Richard Elmore, I am grateful to him for his extensive study. I was greatly informed and influenced by his work, particularly concerning the history of Follow Through.

I completed the paper in April, 1986. Largely through the efforts of Donald Cook and Carl Binder, David Hackett, Executive Director of the Youth Policy Institute, published a special issue of Youth Policy in the summer of 1988 on behavioral instruction. An excerpt from my area paper titled "Project Follow Through: A Story of the Identification and Neglect of Effective Instruction" was included in this issue of Youth Policy. There was a great deal of interest in the article, and I continue to receive requests for reprints.

Although official data-collection for the national evaluation ended in 1976, funding for Follow Through models continued until 1995. Recently, in recognition of the fact that 30 years have passed since its inception in 1967, Effective School Practices published a special issue on Project Follow Through. The 1988 Youth Policy article was revised and published in this collection of articles as "Follow Through: Why Didn't We?"

Thanks to the support of a number of people, in particular Bea Barrett and the late Donald Cook, the paper is now being printed in its entirety as a monograph by the Cambridge Center for Behavioral Studies. When I wrote this paper I was a doctoral student at the University of Florida. I am now an Associate Professor in the College of Education at California State University, Stanislaus - back where I first learned about Follow Through. In the time since I wrote the paper I have spent ten years in colleges of education and in public schools. I was a consultant to the School Effectiveness Model, a Follow Through model sponsored by Washington Research Institute. Although my

willingness of the Cambridge Center for Behavioral Studies (Betsy J. Constantine, Executive Director) to publish the paper in its original form.

Meaningful school reform will not be achieved until we acknowledge that how well students learn is a function of how they are taught. We must identify barriers to the adoption and implementation of effective instructional practices. The neglect of empirical evidence and the resistance to effective methods encountered at virtually every level of the educational system is an important area of study. Our efforts and resources might be well spent investigating the conditions that are conducive to the adoption of empirically validated practices, the conditions necessary to ensure the fidelity of their implementation, and the contingencies that would ensure their sustained implementation in the absence of external support.

Our ability to improve schools depends on our commitment to identifying, analyzing, and disseminating effective instructional methods and our determination to create the schools our children deserve.

Cathy L. Watkins, Ph. D.  
Turlock, CA  
May 15, 1997

**Project Follow Through:**  
**A Case Study of Contingencies**  
**Influencing Instructional Practices**  
**of the Educational Establishment**





## **ABSTRACT**

Project Follow Through, originally conceived in 1967 as a social action program to extend Head Start into the primary grades, became an educational experiment aimed at finding effective methods for educating disadvantaged children. Follow Through, in effect, created a national learning laboratory, providing a unique opportunity to study the effectiveness of a variety of educational methods. The results indicated that the Direct Instruction model and, to a lesser degree, the Behavior Analysis model provided viable solutions to the problem of teaching disadvantaged children. Yet, the results of the Follow Through evaluation have been virtually ignored by the educational establishment. This paper presents a case history of Project Follow Through and examines the factors that have led the educational establishment to ignore teaching methods that are effective in raising the academic achievement of disadvantaged children.

## **PART I: HISTORY OF PROJECT FOLLOW THROUGH**

### **Introduction**

In 1964 with the passage of the Economic Opportunity Act (EOA), President Lyndon Johnson declared a full scale war on poverty. While efforts to alleviate the problems of the poor had historically consisted of some form of income distribution, there was a growing consensus that education would be the antidote to poverty by providing the skills needed to break the existing cycle. This growing interest in education led to the initiation of federally funded compensatory education programs. Compensatory education may be defined as education that is designed to compensate for deficiencies in a person's learning experiences (McDill, McDill, & Sprehe, 1969). The intent of compensatory education programs was to eliminate the discrepancies in achievement between disadvantaged children and their more affluent peers. Perhaps the best known example of a program to develop from this line of reasoning was Head Start. Head Start was conceived of as a comprehensive approach to early childhood intervention aimed at eradicating poverty and was authorized under Title II of the EOA (Zigler & Valentine, 1969).

Head Start began in the summer of 1965 and was immediately popular. However, some educational researchers interpreted data from early evaluations as suggesting that any gains achieved by Head Start participants dissipated when these disadvantaged children entered the primary grades. Thus, it appeared that permanent gains in academic achievement would require extending the program into elementary schools. When the first session of the ninetieth Congress met in February, 1967, President Johnson requested that Congress establish a program to "follow through" on Head Start.

This request resulted in the passage of Public Law 90-92, which authorized Project Follow Through. Follow Through was originally conceived as a social action program that would extend Head Start into the primary grades. However, before the program got underway, circumstances prompted a shift from a social service program to an educational experiment aimed at finding effective methods for educating disadvantaged children. The particular events leading to this change in emphasis will be discussed in a following section. What is important to understand is that, regardless of the particular factors surrounding the decision, the result was the launching of the largest educational experiment ever conducted.

Follow Through provided a unique opportunity to study the effectiveness of a variety of educational methods. Whereas most educational research is conducted in laboratory schools created by universities or research centers, Follow Through, in effect, created a national learning laboratory. There had been no prior attempts to systematically vary instructional methods in order to assess their relative effectiveness.

The evaluation of Project Follow Through was the most costly evaluation study in education ever financed by the federal government. The results of the evaluation indicated that the Direct Instruction model and, to a lesser degree, the Behavior Analysis model provided viable solutions to the problem of teaching disadvantaged children. These two models demonstrated that the technology exists by which the academic achievement of disadvantaged children may be raised to "self-sustaining" levels (Greer,

## *PROJECT FOLLOW THROUGH*

Alessi, Graf, & Lindsley, 1983). The identification of a technology that would make it possible to teach academic skills to children who have traditionally been considered refractory to instruction has enormous implications for comprehensive reform of education policy and practice. Yet, the results of the Follow Through evaluation have been virtually ignored by the educational establishment.

Effective teaching methods exist but are not used to improve education, and it seems imperative to understand why this is the case. It is proposed that the answer may in part be found in the manner in which instructional method is viewed by the various elements of the educational industry. Thus, in order to understand the failure of an industry to use methods that have the greatest potential for increasing student achievement, it is necessary to examine how the various elements of the industry conceive of the interaction of students and teachers—the method of instruction.

The focus of this paper is on how the educational industry approaches educational methods. It is an effort to use Project Follow Through as an opportunity or mechanism for examining instructional method as it is dealt with in all respects by the educational industry at all levels. Follow Through provides an appropriate vehicle for such an analysis because it reveals how the educational industry collectively conceived of, planned, conducted, and interpreted a large scale experiment of instructional methods. Social experimentation of this nature involves intervention in an ongoing system, and studying the implementation of Project Follow Through provides insight into the conditions necessary for implementing educational change. Follow Through thus provides an opportunity for close examination of the influences determining the use of a particular method of teaching.

The paper is organized as follows. Part I presents the history of Follow Through—the genesis of the Follow Through idea, early planning, and the factors leading to the decision to experiment. Part II concerns the actual design and conduct of the experiment. Discussion focuses on the various sources of influence that determined the experimental design. Part III concerns the evaluation. A history of the evaluation effort is presented, followed by a description of the national evaluation and summary of effects. This is followed by the presentation of alternative analyses and interpretations of the data. Part IV documents the neglect of the data by the educational establishment. In Part V each of several components of the education industry is examined for reasons why data of this import have been virtually neglected. Discussion focuses on factors that influence the adoption of teaching methods by the various facets of the educational enterprise. Part VI presents a summary and offers some suggestions about how to solve the instructional problem.

### **Origins**

The impetus for Project Follow Through is commonly attributed to evaluations of Head Start that suggested disadvantaged children enrolled in Head Start classes made substantial academic gains and that these gains had essentially dissipated by the time those students completed their first year in traditional elementary classes (Bissell, 1973). It was asserted by promoters of Follow Through that the Follow Through program was necessary in order to maintain the benefits of Head Start. While it may be comforting to think that the decision to invest many millions of dollars in any federally funded program

## *HISTORY OF PROJECT FOLLOW THROUGH*

is based on solid empirical evidence establishing the need for such a program and indicating the probability of its success, a number of circumstances lead to the conclusion that no such rational decision was possible in the case of Project Follow Through. First, there is evidence that President Johnson already had considered the possibility of vertical expansion of Head Start into the primary grades before there had been any opportunity to evaluate Head Start programs. The very first Head Start programs began in the summer of 1965 and consisted of six to eight week summer sessions. A scant three months after these summer programs were underway, and in the absence of any evaluative data, President Johnson declared that Head Start had been “battle tested . . . and proved worthy” (quoted in Elmore, 1976, p. 115). At that time, Johnson also suggested that plans be developed to enlarge Head Start in order to serve a greater number of children, to extend the summer sessions into year long preschool programs, and to “follow through” on the gains children had presumably made in Head Start after they reached school age.

The first year-long programs started in 1965-66. In January, 1967, when President Johnson presented his formal proposal for Follow Through, only one group of children had graduated from the year-long program. It was too early to obtain follow-up data on these children because they were at that time midway through their first year in public school. Dissipation of the effects of these year-long programs, if such a phenomenon occurred, could not have been measured at that time. In fact, when Follow Through was formally proposed by President Johnson, the only follow-up research available were post hoc studies conducted to try to determine the effects of the summer programs that had been implemented in 1965.

One of these studies in particular (Wolff & Stein, 1966) has been cited as providing evidence of diminution of educational advantages provided by Head Start. The pervasive influence of this particular study was described by Robert Egbert (former Director of Follow Through): “. . . the decision to request a Follow Through program was stimulated by a single follow-up evaluation of children who were enrolled in the first . . . Head Start projects in the summer of 1965” (Egbert, 1973, p. 121). Because the Wolff and Stein study is credited with empirically establishing the need for Follow Through, it is instructive to examine closely just what evidence that study really provided.

To say that the study is methodologically weak is an understatement. While it is only possible to speculate about the specific variables controlling the verbal behavior of the authors when they concluded that effects of Head Start were dissipating by the end of the first year in school, it seems clear that this was not a dispassionate interpretation of obtained data. Briefly summarized, the “design” was to administer posttests (no pretest measures were obtained) to children in New York City six to eight months after their participation in various Head Start summer programs and to compare these scores with a group of children who had not attended Head Start. The difference between scores of the two groups on the Caldwell Pre-School Inventory was found to be insignificant. These results were interpreted by Wolff and Stein as indicating that initial gains the Head Start children made had faded away until their performance was equivalent to that of children who had never had the advantage of this compensatory program.

It does not require a great deal of sophistication in experimental method to realize that the absence of pretest data renders the study without merit. The results provide absolutely no information about the dissipation of gains made as a function of

## *PROJECT FOLLOW THROUGH*

Head Start as there was no evidence that there were any such gains made by those children in the first place. In fact, not only was measurement neglected at the children's entrance into Kindergarten, no assessment had been done prior to the children's initial placement in the Head Start program. The deficiencies in general experimental methodology and analytic procedures prompted Bronfenbrenner (1967) to conclude that it is doubtful that "the study would have been accepted for publication in any established journal" (p. 2).

The point here is not to call into question the professional competence or integrity of Wolff and Stein or to argue that there was no empirical basis for a compensatory program aimed at the primary grades. The point is that if there were data indicating the need for such a program, they were not available at the time the decision was made to extend Head Start into the primary grades. What is important to realize in regard to the Wolff and Stein study is that if, as Egbert (1973) claimed, this single study was the sole basis for the decision to enter into an intervention program of the magnitude of Project Follow Through, there was an astounding lack of sophistication regarding the experimental basis for such a decision.

However, despite Egbert's claim, in light of what was known about the effects of Head Start at the time Follow Through was proposed, it can only be concluded that the decision was based on some other considerations and not on knowledge gained from formal investigation of the effects of Head Start preschool experiences. Because Follow Through was an executive rather than a congressional decision, the most appropriate place to look for alternative explanations is in the various sources of control over the decision making behavior of President Johnson and his representatives at the Office of Economic Opportunity (OEO). Although the decision to initiate a Follow Through program was most likely a function of both conviction and expediency, political advantage, often cited as the primary determinant of Johnson's support for War on Poverty programs in general (Gelfand, 1981), appears to have been the primary influence on the decision.

President Johnson had created the OEO to serve as the strategic headquarters for the War on Poverty. One of the central features of the war on poverty was the notion of maximum community participation. This concept was operationalized by the OEO as the Community Action Program (CAP) which established local community agencies that would be responsible for developing and administering programs to meet the needs of the poor. However, the CAP quickly fell into political disfavor when it resulted in the political mobilization of the poor and the undermining of local government agencies. The political controversy generated by OEO's Community Action Program prompted Congress in the Spring of 1966 to devise legislation that would direct money away from community action agencies and provide funds for specific "national emphasis programs" (Donovan, 1967). Head Start was such a national program. By proposing a program to extend Head Start into the elementary grades, it was possible to capitalize on this already successful program and divert attention from the controversy surrounding the OEO. As one political analyst explained, where the locally sponsored programs had resulted in controversy, "the immense popularity of Head Start, the first national program, stimulated OEO to advance to others" (Levitan, 1969, p. 72).

## HISTORY OF PROJECT FOLLOW THROUGH

Thus, although Johnson's statements indicate that as early as 1965 he was considering enlarging the scope of the federally funded Head Start program to include the primary grades, the formal proposal for Project Follow Through came at a time when a big political success was needed to put OEO back in the good graces of Congress. In addition, Project Follow Through was almost ensured public support by riding on Head Start's coattails, and a popular domestic program would result in much needed political benefits in the troubled times of the mid-sixties.

The announcement to the public that Follow Through "received impetus from the success of Head Start and from a study that showed that many deprived children quickly lost the benefit of Head Start when they began in school" ("U.S. selects 30 school districts," 1967) was misleading. Such statements seem more appropriately characterized as misrepresentations than misinterpretations of the study. Head Start evaluations did not inspire the creation of Follow Through. President Johnson and OEO officials were already of the opinion that it was a necessary program for reasons unrelated to the benefit that might result for disadvantaged children. Mere opinion, however, is often difficult to sell to Congress or to the public, and Wolff and Stein may have provided the evidence necessary to convince Congress to pass this legislation.

The proposal the Johnson administration sent to Congress called for

... a program to be known as "Follow Through" focused primarily on children in kindergarten or elementary school who were previously enrolled in Head Start or similar programs and designed to provide comprehensive services and parent participation activities as described in [the Head Start] authorization, which the Director [of OEO] finds will aid in the continued development of children to their full potential. (quoted in Elmore, 1976, p. 129)

When the program was announced to the public, it was proclaimed that the pilot project would serve about 3,000 children in 25 states and that the program would be expanded in the following year to serve 190,000 children at a cost of \$120 million ("U.S. selects 30 school districts," 1967). Planning and development based on the conception of Follow Through as a large scale social program began long before Congressional approval.

### Planning

Although it has been referred to as the largest and most expensive social experiment ever launched (McDaniels, 1975), Follow Through did not originate as an experiment as much as a comprehensive program that was designed to carry Head Start into the early elementary grades. It is important to remember that what established the need for compensatory education programs was the fact that economically disadvantaged children tend to perform poorly on measures of academic achievement relative to their more affluent peers. It was reasoned that federally funded compensatory education programs would provide these children with an education that would reduce, if not eliminate, the discrepancy between the two groups. "Improving life chances" is frequently cited as the national objective of Project Follow Through. But the philosophy behind compensatory education was that through education the academic skills of disadvantaged and non-disadvantaged children would be made more commensurate. It

## *PROJECT FOLLOW THROUGH*

was this increase in skills that was presumed to increase the opportunities of the poor to break out of the “cycle of poverty.” There are many indications however, that the initial interest in Follow Through was not in finding effective ways to teach the disadvantaged. Instead, emphasis was on gaining political support for the program.

Follow Through was to be authorized under section 222 (a) (2) of the Economic Opportunity Act (EOA) and thus was within the administrative jurisdiction of OEO. One of the first decisions to be made was for OEO to delegate administrative responsibility to the Office of Education (OE) at the Department of Health, Education, and Welfare. Although this transfer of authority was not formally approved until June 1967, early in that year OE began to functionally assume administrative duties.

One of the first planning decisions OE made was to establish a National Follow Through Advisory committee, chaired by Dr. Gordon J. Klopff, Dean of the Faculties of Bank Street College of Education. This committee was made up of fifteen representatives from the area of early childhood development, from academic institutions, and from state or local educational agencies and various government agencies. These representatives, with their disparate histories, provided task force reports on what Follow Through should look like. The advisory committee did not, however, make specific suggestions regarding the actual content of the program or recommendations for evaluation of the effectiveness of such an effort.

The general recommendations made by the advisory committee were that an educational program for the disadvantaged should include early intervention and should be comprehensive in nature. These two “recommendations” represent the consensus of professionals in education and child development. Such global suggestions reflect the influence of developmental psychology on educators and the extent to which these representatives determined instructional method to be of importance. The recommendations of the committee were based “neither on research about what worked educationally with disadvantaged children nor a curiosity about what worked. They were simply the consensus among people with an interest in the program” (Elmore, 1976, p. 144).

The plan at this stage was to begin the operation of a pilot program in 1967-1968 that would provide useful information to be used in designing the greatly expanded program that was anticipated in following years. In June, 1967, pilot sites were selected. Thirty projects were funded at \$85,000 each. However, later that summer, Follow Through administrators responded to “pressure from big-city mayors and congressional delegations” by including ten more sites in the pilot program (Haney, 1977a). This was just the first of a large number of decisions made in acquiescence to the demands of interest groups and while the consequences of such decisions were less critical at this stage, later similar decisions had deleterious effects on the conduct and interpretation of the evaluation. In September of 1967 Follow Through began in the first 30 sites with the additional ten to begin operation the second half of that year.

What was remarkable about the planning stage of the program was the total disregard for instruction. Although Follow Through was intended to provide disadvantaged children with an educational opportunity that would benefit them where traditional education had failed, there were no guidelines for the educational component of the program. The National Follow Through Advisory Committee had promoted a free

## HISTORY OF PROJECT FOLLOW THROUGH

for all with no regulation of instruction. In fact, they contended that the resulting random variation would be preferable. It is difficult to find any logic in this position since presumably the result of random variation would amount to instruction similar to that of regular classrooms not receiving federal funds. Since it had already been well established that this population did not benefit from the instructional approaches found in traditional classrooms, there was no basis for believing that they would profit from the same instruction in the context of Follow Through. It must be concluded that the educational community, represented by the National Follow Through Advisory Committee, did not recognize the role of instructional methods in education.

### The Decision to Experiment

#### Budget Cuts

All planning during the pilot phase proceeded on the assumption that \$120 million appropriated by the Office of Economic Opportunity would be available to cover the cost of the pilot phase as well as the enormous expansion of the program that was planned for the following academic year (Elmore, 1976). However, during the fall of 1967 it became increasingly clear that there would be drastic cuts in the OEO budget, as Congress tried to curtail domestic spending. This placed Follow Through in jeopardy because it meant that the broad scale expansion of the program that had been proposed would not be economically feasible. Predicting that drastic cutbacks in OEO's budget would result in the elimination of Follow Through, the program's administrators began to cast about for a reason for the program's continued existence.

The idea of shifting the orientation of the program to research was initially proposed by John Hughes, Director of the Division of Compensatory Education, and Robert Egbert, Director of Follow Through, in a memo to the Associate Commissioner of Elementary and Secondary Education, which read in part:

Should the worst happen (a final \$14 million appropriation for Follow Through), Bob Egbert and I feel that the remaining funds . . . should be used to fund some research-oriented innovative pilots in thirty or so hand-picked communities. (quoted in Elmore, 1976, p. 155)

In December, 1967, when Congress passed the amendments to the Economic Opportunity Act that authorized Follow Through, OEO's budget had been reduced to \$1.77 billion. OEO allotted \$15 million to the Follow Through program and justified the continued existence of the program on the basis that it would function as a research oriented effort designed to answer questions about the effectiveness of compensatory education programs.

Thus, Follow Through was converted from a large scale comprehensive social program to an educational experiment. As Egbert (1974) later explained, the funding situation made it seem "sensible to change Follow Through's primary purpose from 'service to children' to 'finding out what works'" (p. 122). However, Congress had authorized the program in accordance with President Johnson's original proposal, which clearly called for a service program. The proposal made no mention of experimentation or research, nor was any change made in the wording of the legislation following the



## PROJECT FOLLOW THROUGH

reincarnation of Follow Through as an experiment. This failure to communicate to Congress the change in orientation and to alter the authorizing legislation served as a source of confusion and controversy that continually surrounded Follow Through (Egbert, 1974).

Although the decision to convert Follow Through to an educational experiment provided an immediate solution to the problem raised by the budget cut, the idea was not spontaneously generated. The seeds from which the notion of planned experimentation grew had been planted by policy analysts who had made federal policy makers increasingly aware of the need for better program evaluation.

### White House Task Force

In October, 1965, the Bureau of the Budget established the Planning, Programming, Budgeting System (PPBS) in all federal agencies and departments. The intent of the PPBS was to provide data on program effectiveness in order to determine the most effective allocation of agency funds. The concept of PPBS was put into practice by creating offices of planning and evaluation in federal agencies that administered domestic programs. For example, at HEW the office of the Assistant Secretary for Planning and Evaluation (ASPE) was established; within HEW at the Office of Education the office of Program Planning and Evaluation (OPPE) was created.

These offices were staffed by policy analysts whose task it was to evaluate programs funded by the federal government and to establish a data base that policy makers could then use in the formation of public programs. While the effects of PPBS are not agreed upon by all policy analysts, some (e.g., Rivlin, 1971) argue that it resulted in focusing attention on the effects of public programs funded by the federal government. However, Elmore (1976) concluded that the real and enduring effect of PPBS was in the “emergence of policy analysts as a political force in public decision making and with them the notion of rational choice among discrete policy alternatives based on the outcomes of those alternatives” (p. 38). Elmore (1976) explained that the central issue of rational choice is

... not simply judging whether programs work, but what kinds of alternatives within a given program work best. The ideal evaluation, under the model of rational choice, is one that provides an array of program alternatives. The exercise of rational choice in public decisionmaking consists of *using evaluation results to reform programs by eliminating ineffective alternatives and emphasizing effective ones* [emphasis added]. (p. 26)

In September, 1967, William Gorham, HEW Assistant Secretary for Planning and Evaluation was requested by Special Assistant to the President, Joseph Califano, to review HEW's programs for youth and, based on the findings, to make recommendations for legislative and administrative decisions. In response to this request, Gorham organized The White House Task Force on Child Development made up of federal officials and academicians. With respect to what was known about the academic deficits of disadvantaged youth the Task Force stated:

... studies have amply documented the inadequate educational performance of low-income and minority group children. These children start school at a disadvantage. As

## HISTORY OF PROJECT FOLLOW THROUGH

they pass through the school system they fall more and more years behind the majority in educational achievement measured by standard tests. (Gorham, 1967, p. 18)

In summarizing what was known about how to alleviate this problem the Task Force concluded:

The one thing that stands out in all our efforts to analyze education programs is our lack of knowledge about effective ways of raising the educational attainment of students. Research techniques must be combined with experimentation and innovation to throw light on this problem, and *widespread dissemination of the most effective methods must be undertaken* [emphasis added]. (Gorham, 1967, p. 19)

The Task Force was fully aware of the impending budget cuts and the uncertainty of Follow Through and suggested that the new program would be an appropriate means by which to accomplish the objective of delineating and disseminating effective teaching methods. Specifically, they recommended (a) a major effort to use federal funds to learn how to make education of disadvantaged children more effective, and (b) use of the Follow Through program to evaluate major variations in compensatory education programs in the early elementary grades (Gorham, 1967, pp. 158-163).

The report prepared by the White House Task Force on Children included suggestions from two of its members (David Cohen, from the Harvard-MIT Joint Center for Urban Studies, and Urie Bronfenbrenner, from Cornell University) regarding appropriate components for systematic variation. Four of the variables recommended were pupil to instructional staff ratio, racial and socio-economic mix, age and training of teachers and aides, and type of parent involvement. The suggestion of these variables indicates that the Task Force members collectively endorsed a practice common in educational research in which fixed or sociological variables are correlated with student outcomes (Greer, 1982).

This strategy, using similar types of variables, had historically been used in the evaluation of other compensatory education programs. For example, in 1965 HEW had commissioned TEMPO, a division of General Electric, to evaluate the effects of Title I programs. TEMPO concluded that no inferences could be made about the effectiveness of Title I because the program itself was not properly designed to yield such data. Failures in educational research such as this were, in some measure, responsible for the growing support that the concept of planned variation generated among policy analysts. Yet, when recommending important sources of variation, the Task Force suggested fixed and sociological variables.

Cohen (cited in Elmore, 1976) had suggested that in deciding what to vary systematically priority should be given to those variables that could most readily be controlled, measured, and replicated. Although the variables suggested met those requirements, they were not necessarily directly related to student outcomes.

Elmore (1976) postulated that some variables are sufficiently difficult to describe and measure that they are of limited usefulness as sources of variation. However, it seems unlikely that our understanding of the variables related to student learning will be advanced by relying only on those variables that are easy to define categorically. Convenience is probably a poor criterion for experimental decision making.

## *PROJECT FOLLOW THROUGH*

A more appropriate place to search for factors related to student outcomes is in the exchange between student and teacher, or the instructional method. Only one variable included in the Task Force recommendations came close to addressing this relationship; Urie Bronfenbrenner proposed that it would be important to systematically vary “curriculum content” when trying to determine what works best in compensatory education. Curriculum is an all encompassing term used alternatively to describe a general philosophy of schooling and the more specific characteristics such as teaching method and materials. Although curriculum became the major focus of the Follow Through experiment, failure of the Follow Through administration to define curriculum in an operational manner resulted in an evaluation that was perhaps less precise than might have otherwise been achieved.

While policy analysts at HEW had promoted a model of rational choice for decision making for several years, no program had yet been developed using this model. The budget reductions and precarious position of Follow Through afforded policy analysts an opportunity to muster support for the model of rational choice. At the same time, the notion of planned experimental variation provided Follow Through with a reason to be.

It is not clear to what extent the Task Force recommendations influenced the decision to redesign Follow Through around the concept of planned experimentation. Robert Egbert, Follow Through Director, had expressed an interest in experimentation during early planning stages, before his appointment as Director (Elmore, 1976). Yet, there were certain events that occurred and decisions that were made that suggest that experimentation was proposed not because of any true desire to identify appropriate educational strategies but rather as a means to an end, that end being the immediate survival of Follow Through and the eventual enlargement of the program into a large scale program. It is clear that Egbert’s interest was not so much in identifying effective methods of instruction as it was in generating support for the program in order to obtain the financial support needed to expand its scope. Egbert (1974) was aware that unless “substantial improvements [in student outcomes] were manifest, the really massive increases in spending that would be required [to expand the program] could not be justified” (p. 128). With this as an objective, it became important to create the largest constituency possible. This objective however, was often at cross purposes with experimental requirements. Egbert made many decisions that were intended to placate program supporters, but which had the additional consequence of compromising the integrity of the experiment.

*EXPERIMENTAL DESIGN*

## **PART II: EXPERIMENTAL DESIGN**

### **Planned Variation**

The design of the Follow Through experiment has come to be known as planned variation. This approach was based on the notion that a variety of curricula and instructional methods consistent with alternative philosophies of learning could be designed, implemented, and evaluated (McCandless & Evans, 1973). Recall, however, that Follow Through was not legislated as an educational experiment but as a large scale social program. Although a substantial reduction in OEO's budget had occasioned a shift in Follow Through's orientation from service to research, this change was never reflected in legislation. Follow Through's enabling legislation, the Equal Opportunity Act, imposed a number of service program requirements that were not compatible with the requirements of experimental design (Egbert, 1974). For example, EOA required that local community action agencies and parents participate in project planning. The goals of agencies and parents however, are likely to be disparate from those of experimentation. Their suggestions and demands had to be considered, however, at the risk of being found in violation of the statute. As a consequence, the concerns of these parties had to be given preference over the methodological requirements of experimentation. In addition, the EOA stipulated that the program provide medical, dental, social, and psychological services, as well as nutritional programs. As will be discussed shortly, provision of such social services were particularly problematic for the design of a controlled experiment.

All EOA service programs were subject to the requirements mentioned and Follow Through, at least according to the legislation that authorized it, was such a program. Thus, Follow Through administrators had to either comply with the EOA regulations or obtain a waiver of the requirements. However, as Elmore (1976) pointed out, it is likely such a waiver would have involved a Congressional amendment formally changing Follow Through from a service program to an experiment. Moreover, it would have required Follow Through administrators to inform Congress that the change in orientation they were requesting had already, for all intents and purposes, been made. Had Follow Through administrators been able to secure a legislative amendment waiving the EOA stipulations they could have designed Follow Through in accordance with the requirements of experimentation independent of social and political interests. In fact, much of the controversy surrounding Follow Through has been occasioned by conflicting opinions about the purpose of Project Follow Through. This source of confusion could have been eliminated by a legislative amendment and by the administrators declaring unequivocally what the goals and objectives of the experiment were. However, the goals and objectives of Follow Through administrators themselves conflicted with the requirements of experimentation.

Despite the fact that Egbert (1974) has acknowledged that he was aware of the constraints the EOA requirements would impose on experimental design, he elected not to seek exemption from the requirements. Furthermore, he made no apparent effort to inform interested parties that Follow Through was being conducted as an experiment. Such behavior is clearly inconsistent with designing an experiment intended to investigate the relationship between instructional variables and student outcomes.

## *EXPERIMENTAL DESIGN*

However, there is evidence that the experiment was never considered by Follow Through administrators to be a means of gaining knowledge about effective instructional methods, but rather, was a means to a different end, namely that of expansion of the program into its original envisioned status as a large scale social program. This motivation is evident in Egbert's (1971) statement that OEO, HEW, OE, and Bureau of the Budget officials all agreed that "Follow Through – for the time being – should be an experimental program designed to produce information which would be useful when the program is expanded to nationwide service proportions" (p. 7). This is a very important point because decisions that determined the design of the experiment were based not on their effect on the design and conduct of the experiment and evaluation, but rather, on the degree to which they would enhance the probability of increased funding the following year.

The decision not to pursue a Congressional waiver of EOA requisites is a case in point. While a Congressional amendment specifically limiting the scope of Follow Through to experimentation would have freed administrators from the constraints of EOA, it would also have made it difficult to enlarge the program to service proportions at a later date. The administrators therefore elected to attempt to conduct the experiment under the constraints imposed by existing legislation, despite the resulting compromises in experimental rigor. By doing so the necessary legislation would already be in effect if and when the program was expanded to its full service program status.

There were other potential long term benefits to be gained from including comprehensive services in Follow Through. For example, there was a decided advantage to maintaining a close association with Head Start. The original vision of Follow Through was one of direct vertical expansion of Head Start into the elementary grades. This was, in fact, the intent of the authorizing legislation. Congressional approval had been based, at least in part, on the "evidence" presented by OEO that gains made as a function of Head Start were dissipating without further similar services in the primary grades. The more closely Follow Through resembled the enormously popular Head Start, the more it could capitalize on the success of the preschool program (Elmore, 1976). Thus, the reasoning went, since Head Start included comprehensive services, Follow Through should also include similar services so that both social and educational components would be ready for expansion when the time came.

A second benefit of including comprehensive services was the continued support of professionals in the areas of child development and compensatory education (Elmore, 1976). Experts in these areas had endorsed comprehensive services as necessary and sufficient for "improving the life chances" of the disadvantaged. Witness the recommendations of the National Follow Through Advisory Committee. That the program should include comprehensive services was as detailed a prescription that this group of professionals was able to provide. For Follow Through administrators to have unilaterally declared that provision of comprehensive services be waived in order to obtain the experimental control necessary to systematically vary instruction could potentially have been construed as undermining the opinion of these experts. Egbert apparently felt that the potential benefit to be gained from professional support far outweighed the clarity of evaluation afforded by a stringent methodological design.

The failure to obtain a waiver of EOA requirements and to explicitly state the experimental nature of the program resulted in a number of choices, compromises, and

## *PROJECT FOLLOW THROUGH*

commitments that were made with the knowledge that they would have deleterious effects on the design of the experiment and subsequent interpretation of the data. Some of the administrative decisions made during the planning of the experiment will be discussed as well as the implications of such decisions for the evaluation.

As already mentioned, the inclusion of comprehensive services mandated by EOA posed a special problem. It was decided very early that it was beyond the realm of political possibility to impose systematic variation of these comprehensive services; for example, to include dental services at one project but withhold those services from a different project. An alternative to experimental manipulation of these services would have been to arrange for their uniform provision at all sites. By making the delivery of these services constant, any influence they had collectively on outcome measures could be controlled. This tactic was not pursued either. Instead, it was left to local projects to determine how these services would be provided. In other words, these services were not subject to experimental control but introduced an additional source of unsystematic variation into each site. Thus, there were in effect two components to Follow Through: an unsystematically varied social services component and, overlaid on top of this, a systematically varied educational component. The planned variation in educational approaches was the independent variable of the Follow Through experiment. The presence of the additional services however, could potentially influence the outcome variables as well. Although there is little or no evidence to support the notion that comprehensive services, either independently or in any combination, have a direct relation with student achievement, the possibility exists that such a relationship obtains. Because the effects of these social services have not been determined, the extent to which they affected outcome measures, if any, is an unknown. If such an effect does exist, then it will necessarily be confounded with the effect of the experimental variable, instruction. In addition, as Elmore (1975) pointed out, if these services have no effect on education, then it makes no sense to promote them as essential to an educational program.

Comprehensive services were not the only confounding variable administrators introduced into the Follow Through experiment. Once having decided to vary only the educational component, the question of what these educational approaches would be, where they would come from, and how they would be put in place in the public schools had to be addressed.

### **The Sponsored Model Approach**

While there was a general consensus among OEO, HEW, OE, and the Bureau of the Budget to alter the focus of Follow Through to research, it was not clear how the concept of planned variation would be operationalized. There was certainly no precedent for such an approach and no guidelines for implementing and administering such an effort. One certainty was that having federal officials mandate the adoption of a particular educational model by a school district was antithetical to the notion of local control that is the cornerstone of the American educational system. Robert Egbert (Director) and Richard Snyder (Follow Through's Chief of Evaluation) devised a mechanism of implementation that would enable a variety of educational methods to be put into effect in local school districts while avoiding the appearance of unwarranted federal intervention. The plan, based on the novel concept of model sponsors, was that OE would

## *EXPERIMENTAL DESIGN*

contract with a number of developers of educational approaches. These developers would then act as sponsors of their approach and work cooperatively with local school districts to implement the model in Follow Through classrooms. While Follow Through's division of program management would monitor the relationship between sponsor and district, federal officials would not directly intervene in model implementation.

The task proposed for the Follow Through sponsors, designing a complete curriculum for the entire school day's activities, had never been attempted in educational reform (Hodges, Sheehan, and Carter, 1979). In short, the sponsors would be responsible for translating their particular approach to education into practice. This would require the selection or development of instructional materials as well as training teachers in the instructional method upon which the model was based.

While sponsorship seemed to offer a reasonable means of implementing diverse instructional approaches in the public schools, it was a novel approach to intervention. Numerous decisions had to be made in order to convert the concept of sponsorship from an administrative idea into an educational experiment. The decision to use a sponsored model approach was made in December, 1967. Follow Through was scheduled to begin operation the following September. This left only eight months to finalize the design of the largest educational experiment in history, and everything that needed to be done had never been done before. It was imperative that sponsors and sites be selected and paired immediately in order to allow them as much time as possible to prepare for the project.

### **Selection of Sponsors**

The selection of sponsors began in January, 1968 during two meetings that took place in Washington, DC. Richard Snyder, Chief of Evaluation, and his assistant, Frieda Denenmark, invited about two dozen individuals or groups to participate in these meetings. It is not clear what criteria were used in selecting participants; invitations were apparently extended to any group conducting research on instructional methods.

The invited participants, representing universities, research laboratories, and privately or federally funded local projects, were scheduled to give a brief presentation of their programs followed by a discussion of their approaches with Egbert, Snyder, and other members of the Follow Through staff. Consultants to OE as well as various OEO and ASPE representatives also attended these meetings.

The intent of the experiment was to evaluate educational approaches in order to determine the most effective means of educating children. In order to make a meaningful comparison, the dimensions along which the treatments would vary had to be specified. A reasonable strategy might have been to determine the outcome of interest, then to specify a number of dimensions that might be expected to be related to the outcome measures, and to introduce variation along these dimensions. However, neither the outcomes nor the dimensions of variations were specified in advance. It was assumed that, as the developers presented their various approaches, important differences between the programs would be revealed. These variations would then form the basis for the evaluation. The sponsors, however, were so diverse in their views that they could not even agree on the goal of education, let alone on the important sources of variation.

The Transcript of Proceedings (cited in Elmore, 1976, pp. 202-207) from the meetings indicate that some potential sponsors suggested an alternative to determining



## PROJECT FOLLOW THROUGH

the dimensions of variation. Sigfried Engelmann (Direct Instruction Model), and Ira Gordon (Florida Parent Education Model), were two program developers who argued for a precise definition of the individual components of each program. This definition of treatment variables would itself identify differences between models. It also would allow for an analysis of effects that would provide information about what components were more or less effective, and under what conditions. The sponsors promoting this approach to specification of treatment were outnumbered. The majority of potential sponsors had no interest in such operational definitions and were not in a position to do so in any event because they were, with few exceptions, unable to identify critical components of their programs. In fact, Elmore (1976) observed that rather than developing a set of teaching procedures based on theory or basic learning research, most developers had “simply taken a novel idea – open classrooms, parent education, community control – and improvised a program around it . . .” (p. 198).

The issue of specification of treatment was not resolved at this meeting. In fact, it was never adequately resolved. The administrators dealt with the inability to precisely define program variations by ignoring it. Since they had no specific criteria for selecting sponsors, they elected to extend an invitation to submit a formal proposal to any developer who had any semblance of a program. Of the 18 developers issued this invitation, 16 subsequently presented proposals. Twelve of these proposals were accepted. According to one Follow Through staff member, the basis for rejection was simply “. . . our judgment that the applicant simply didn’t have a program” (quoted in Elmore, 1976, p. 208).

The twelve sponsors began implementing their models in various communities throughout the country in September 1968. These sponsors had been recruited because they represented a wide variety of instructional approaches. They were not, however, to be the only Follow Through sponsors. Ten additional sponsors were added over the following three year period. These new models were included not because they offered unique approaches to compensatory education, but because they offered the possibility of enlarging the Follow Through constituency (Elmore, 1976). This second group of sponsors included state education agencies, minority interest groups, and profit-making firms. It should be understood that the participation of these additional sponsors was not solicited. They were added in response to political pressure they exerted on federal administrators. Because they did not represent any particular instructional method, this added yet another source of difficulty in defining dimensions of variation.

### **The Follow Through Models**

Because of various criteria for inclusion in the evaluation sample, not all of the eventual 22 models were included in the national evaluation. The models that were part of the evaluation sample will be described in more detail in a later section of this paper. The following table lists each model and its sponsor.

	<i>Model</i>	<i>Sponsor</i>
1	Tucson Early Education Model	University of Arizona

*EXPERIMENTAL DESIGN*

	<i><b>Model</b></i>	<i><b>Sponsor</b></i>
2	Cognitively Oriented Curriculum Model	High/Scope Research Foundation
3	Direct Instruction Model	University of Oregon
4	Individualized Early Learning Program	University of Pittsburgh
5	Language Development (Bilingual) Approach	Southwest Educational Development Laboratory
6	Mathamagenic Activities Program	University of Georgia
7	Responsive Education Program	Far West Laboratories
8	Bank Street Model	Bank Street College of Education
9	Behavior Analysis	University of Kansas
10	Florida Parent Education	University of Florida
11	Interdependent Learning Model	New York University
12	EDC Open Education Program	Education Development Center
13	Hampton Institute Nongraded Model	Hampton Institute
14	Culturally Linguistic Approach	Northeastern Illinois University
15	Parent Supported Application of the Behavior Oriented Prescriptive Teaching Approach	Georgia State University
16	Home-School Partnership Model: Motivational Approach	Southern University and A.& M. College
17	Role Trade Model	Western Behavioral Science Institute
18	New School Approach	University of North Dakota
19	Culturally Democratic Learning Environments	University of California at Riverside
20	Parent Implementation Approach	AFRAM Associates
21	California Process Model	California State Department of Education
22	Responsive Environments Early Childhood Model	Responsive Environments Corporation

## *PROJECT FOLLOW THROUGH*

### **Selection of Sites**

The selection of sites progressed synchronously with sponsor selection. The manner in which school districts were selected showed total disregard for the experimental paradigm. Rather than employing a stratified random sampling plan which would have resulted in the representative distribution of relevant characteristics of school districts, administrators selected sites opportunistically. In December, 1967, The Office of Education (cited in Elmore, 1976) issued a memo requesting that Chief State School Officers and OEO Technical Assistance Officers nominate school districts as potential Follow Through sites. Included in this memo were the qualifications required for consideration as a participant school district. The memo specified that a district should be capable of beginning a comprehensive services program before the start of the school year, be willing to participate in the planned variation experiment, and have a good working relationship with the local community action agency. The states responded to OE's request by nominating 225 school districts.

In January, 1968, a group of OEO and OE representatives, Follow Through staff members, and outside consultants met to review the nominees and designated 51 school districts as Follow Through sites. Site selection had nothing to do with random or representative sampling for the purposes of experimentation, but was based on administrative considerations. In other words, sites were selected that were believed to have a high probability of successfully implementing the program.

Although this practice has been criticized for introducing experimental bias, limiting the experimental sites to those having particular features that allow for reliable implementation of treatment does not, in and of itself, pose a threat to external validity. Specification of the selection criteria may be conceived of as specification of the experimental environment. The results of an experiment must be interpreted as a relationship between the independent variable and the outcome. If the experiment is contaminated by influences other than the variables of interest, it is not possible to attribute outcome solely to treatment. For this reason experimentation typically involves arranging the introduction of the independent variable in such a way that it will not be distorted.

To the extent that the educational approaches in Follow Through were faithfully implemented, it would be possible to draw conclusions about their relative effectiveness. In the absence of pure implementation, outcome could not as confidently be attributed to the model of teaching. Thus, it makes practical sense to create or select for experimental situations in which the likelihood of sound implementation is increased. Although contrived, highly controlled experimental environments result in greater clarity of the relationship between treatment and outcome.

Uncertainty exists about the extent to which school districts were matched in their ability to execute the program. Accurate implementation is essential in order to attribute differences between models to the model and not to differences in the extent to which the actual practice at a given site resembled the model. This issue has been raised in almost every discussion of the Follow Through experiment and evaluation and will be addressed in greater detail in a later section. For now, the important point concerning the selection process is that selecting districts on the basis of a particular set of characteristics would not have necessarily introduced bias; it would have specified the particular

## *EXPERIMENTAL DESIGN*

experimental conditions. Consequently, the relative effectiveness of properly implemented programs could be determined. How to arrange for the conditions necessary for implementation is a different issue and one that is more relevant for policy decisions than for inference of experimental effects. Despite the shortcomings of the selection procedure, the process was repeated in successive years; 60 additional sites were added for the 1969-70 school year, and 12 more sites were included in 1970-71.

### **Pairing of Sites and Sponsors**

Once the players had been determined, the next step was to create teams. This occurred in February, 1968 during a four day conference held in Kansas City. Once again, experimental convention was ignored. No attempt was made to randomly assign sponsors. Instead, the pairing of sites and sponsors was governed by social and political contingencies. In the American education system, decisions are made primarily at the local level. When federal agencies become involved in state and local education, it is with the approval of state and local officials. In comparison to Follow Through, most federally administered programs require little from local districts other than their agreement to provide certain services (usually social) in exchange for monetary compensation. Even when participating in federally funded projects, school districts operate with maximum autonomy; the manner in which federal funds are spent is largely determined by the individual district. Never had a federal agency become as involved in the actual operation of public schools as they would in Follow Through. There was no precedent for permitting an outside agency to intervene in local schools and redesign the curriculum. There was, therefore, no way to predict how people at the local level would respond under these circumstances.

Although the sponsored model approach was conceived of as a way to minimize the appearance of government control, there was no assurance that the plan would be universally met with approval. There was evidence that at least some individuals viewed the plan as undesirable and unwarranted federal intervention. For example, one former member of the HEW staff (cited in Elmore, 1976) recalled that an Associate Commissioner in HEW had expressed his disapproval of the plan and declared his intention to mobilize state and local interest groups against it. In addition, the plan was presented to school districts as a service; it was implied that sponsors would offer, not impose, instructional assistance to local school districts.

Every effort was made to increase the probability that local sites would cooperate in implementing one of the instructional models. One way to increase the probability of cooperation was to create a situation in which all of the local interest groups would become stakeholders in the program. In order to accomplish this administrators invited representatives of all conceivable state and local interest groups to participate in the pairing of sites and sponsors. After hearing a very brief description of each model by their respective sponsors, local representatives selected the model they believed to be most compatible with the goals and interests of their district. In short, the requirements of sound experimental design were disregarded entirely in deference to the preference of local representatives.

Follow Through administrators offered two reasons for pairing sites and sponsors in this manner. The first has already been mentioned. It was expected that districts would

## *PROJECT FOLLOW THROUGH*

make a greater effort to implement a program that they had personally selected based on the degree to which they perceived that approach as reflecting their own convictions about education. The second reason was that because instructional method is always determined by local choice, not mandated by the government, the experiment would more closely represent the conditions of the real world if selection were based upon local discretion.

The pairing procedure resulted in the inequitable distribution of sponsors across school districts. Sponsors differed in the number of sites at which their model was implemented. The number of sites associated with a particular sponsor ranges from one (e.g., Culturally Democratic Learning Environments Model; Responsive Environments Corporation Early Childhood Model) to twenty (e.g., Tucson Early Education Model; Direct Instruction Model). The sponsor's sites differed in ethnicity as well. The Bank Street Model was implemented in sites that are primarily black, while most students in the Individualized Early Learning Program were Caucasian. The Language Development (Bilingual) Education Approach was designed specifically for and implemented in Spanish speaking communities. In addition, geographic regions were not equally represented among models. For example, the majority of sites served by the Florida Parent Education Model are in the South while the Responsive Education Program has been implemented largely in the North.

A final difference was in the grade level at which program implementation began. Follow Through was intended to serve children beginning in their first year in public school and continuing through the third grade. Whether a child entered the public school system in kindergarten or first grade was determined by whether the child lived in an area that had state-supported kindergartens. Because of these differences in entry level of students, students differed in the number of years of Follow Through participation. The majority of sites where children entered in first grade, as opposed to kindergarten, were in the rural south.

These model distribution problems resulted in the loss of many comparison opportunities (McDaniels, 1975). For example, it would be possible to determine the differential impact of a particular educational model in urban and rural settings only if that model were implemented in both types of settings. Furthermore, the possibility of comparing different models in terms of their relative effectiveness in urban and rural settings is extremely limited since, with few exceptions, models were not implemented in both urban and rural settings.

### **Comparison Groups**

It was the intent of the administrators that the basic evaluation strategy would be to compare the performance of students in the Follow Through models to that of similar students in traditional (non-Follow Through) classrooms. The possibility of random assignment of students to Follow Through or comparison classes was never raised during the planning and selection meetings. Because no plan existed for random assignment of students, assignment of students to Follow Through classrooms was done by local school personnel.

The entire process of selection and pairing of sites and sponsors as well as assignment of students to treatment classrooms had been thoroughly subjective. While

## *EXPERIMENTAL DESIGN*

these decisions constituted the design of the evaluation, all of these events took place prior to the selection of an evaluation contractor. An evaluation contract was awarded to Stanford Research Institute (SRI) only three months prior to the beginning of the school year. Entry level data needed to be obtained in the fall for both Follow Through and comparison students. Consequently, one of the first tasks facing SRI was the identification of comparison groups. Given the lack of random assignment, SRI attempted to construct comparison groups based on their perceived similarity with Follow Through students. However, it proved difficult to identify similar students who had not been assigned to Follow Through classrooms. Because an effort was made by districts to place the most disadvantaged children in Follow Through, comparison groups often turned out to be inadequately matched on socioeconomic variables. In addition, because Follow Through classes were largely populated by Head Start graduates, they differed from comparison classes in terms of preschool experience. In some sites all children meeting eligibility requirements were placed in Follow Through classes and comparison students had to be found in neighboring districts.

### **Summary**

The methodological weaknesses from which the Follow Through experiment suffers are not infrequently found in social research. In fact, such compromises in experimental precision have become so commonplace as to be accepted as inevitable. Yet, it is not necessary that social researchers employ inadequate design and control tactics.

In the case of Follow Through, the early decision to accept the statutory requirements of EOA resulted in a loss of absolute authority by those in charge of conducting the experiment. As previously discussed, the decision was largely influenced by the expectation that Follow Through would be expanded to a large scale social program. This decision, and many subsequent ones, were intended to engender a coalition of supporters for the program. The fact that Follow Through existed for nearly three decades almost solely on the basis of public support attests to the fact that this strategy was effective in creating a strong constituency. However, it was not without cost.

The experimental design of Follow Through was not determined by the empirical question, but rather by the consensus of many individuals including parent and community action representatives. The interests of these parties were seldom congruent with the conditions necessary for experimentation. In fact, many of these people were never informed of, or never accepted, the transformation of Follow Through from a service oriented program to a research project intended to provide information about effective teaching methods. Consequently, decisions were made regarding the conduct of the program that were more appropriate for a social service program than for an experimental effort aimed at obtaining the most precise estimate of intervention effects possible.

Even if those in charge of Follow Through could have obtained control over the planning of the program it is questionable that the design would have been significantly different. The administrators were not scientists primarily interested in identifying the components necessary for effective instruction. The success of the program, from their perspective, was to be measured not by the knowledge acquired but by the extent to

## *PROJECT FOLLOW THROUGH*

which the program's operation could be realized and maintained. Attainment of these goals required public and Congressional support. Consequently, the behavior of the program planners was never controlled by the empirical question as much as by social and political considerations. In short, although adequate sampling procedures could have been used and sponsor and student assignment could have been randomized, social and political contingencies militated against such practices. And because a strong coalition depended on social and political support, the administrators were susceptible to control by these sources.

It should not be assumed however, that stringent methodology would have been the order of the day even if empirical knowledge had been the primary goal of the administrators. The methodological compromises common in social and educational research are often considered to be not only unavoidable, but desirable. It has been argued that highly controlled studies yield causal relationships that do not necessarily generalize to field settings (Campbell & Stanley, 1966). With respect to Follow Through, Elmore (1976) suggested that policy makers are more interested in how educational models perform under close approximation of day-to-day conditions than under tightly controlled contrived conditions. However, the pursuit by educational researchers of identifying educational models that will be effective under "day-to-day" conditions seems misguided. The likelihood must be considered that existing conditions preclude the implementation of effective educational methods. Much knowledge was gained by the Follow Through sponsors about the requirements for program implementation and maintenance. The conditions under which models were implemented were very contrived, but not very tightly controlled. To the extent that local conditions determine the extent to which instructional methods will be implemented, these conditions must be carefully analyzed and altered.

The consequences of the lack of attention paid to methodological issues were not fully realized until the evaluation and interpretation of the results. The inattention to design considerations by Follow Through administrators has affected the precision with which effects can be measured. Consequently, the results of the experiment must be cautiously interpreted.

## PART III: EVALUATION

### The National Evaluation

#### A Select History

*Stanford Research Institute.* The contract for the evaluation of Follow Through was initially awarded in 1968 to Stanford Research Institute (SRI). The evaluation has been, from the very outset, the subject of a great deal of controversy. Early evaluations were characterized by a lack of clarity of purpose. Administrative ambiguity resulted in conflicting perceptions among various interest groups. Shortly after the collection of baseline data had been completed, Follow Through administrators organized a meeting in Atlanta, Georgia, that was attended by Follow Through administrative staff, model sponsors, OEO representatives, personnel from local Follow Through projects, and general consultants. Krulee (1973) provided a detailed account of this meeting that reveals the extent of misunderstanding generated by the administrative ambiguity about the purpose of the Follow Through experiment. It was clear from the nature of the complaints of general consultants and local project personnel that, as far as they were concerned, Follow Through was, and would continue to be, a community action program. The dissatisfactions expressed by representatives of local interests suggest that they were either completely uninformed of, or entirely unconcerned about, the entire concept of a planned variation experiment. For example, local personnel argued that sponsors had too much control over program content and maintained that substantive curriculum decisions should be made by parents and community representatives. There were charges that the evaluation was too limited and that educational outcome was emphasized at the expense of social and community change. In addition, it was charged that minority groups were not adequately represented by Follow Through.

The effects of this meeting are important because they illustrate the degree to which decisions affecting the conduct of the experiment and the evaluation were influenced by pressure from special interest groups rather than by the experimental question. Because there had never been any clearly established goals for Follow Through, and because of the motivation to build a strong coalition, the interests of all parties were considered and, more often than not, acted upon. For example, pursuant to the Atlanta meeting, seven additional models were added to the list of educational approaches. These new models were sponsored by various state educational agencies, private corporations, and minority colleges and were invited to participate not because they represented an important variation in education approach, but because they represented “three different groups not included in the first set of sponsors” (Egbert, 1971, p. 13). In other words, they were included because they would expand the breadth of the constituency.

Also, in response to the demands made at the Atlanta meeting, SRI promised to increase the effort to measure institutional change. Thus, instead of conducting the experiment and the evaluation in accordance with a predetermined experimental design, changes occurred that increased the size of the experiment and the complexity and focus of the evaluation. Local project personnel and general consultants were controlling the evaluation. Experimental control was sacrificed and with it the opportunity to precisely measure treatment effects.



## *PROJECT FOLLOW THROUGH*

The evaluation became the source of much discontent again in 1969 when SRI representatives met with Follow Through sponsors and administrators to discuss plans for the 1969-70 evaluation. SRI had proposed the use of a collection of standardized intelligence and achievement tests commonly used in educational evaluation to measure educationally important skills. It should be clear that the entire concept of compensatory education was based on the observation that disadvantaged children have failed to benefit from traditional educational practices as measured by standardized tests such as those SRI suggested as appropriate measures of model effectiveness. If the entire program was aimed at finding ways to enhance the educational performance of disadvantaged children, and if poor educational performance had been measured by certain standardized tests, then it made sense to employ the same or similar measures to determine the extent to which academic deficits were remediated. SRI was attacked, however, for its selection of outcome measures on the grounds that the instruments comprising the test battery emphasized basic academic skills. Clearly, not all sponsors saw teaching academic skills as the goal of education, and they protested that measurement of academic skills alone would result in a better showing by models that focused on teaching those skills. How could it be otherwise?

In addition, the emphasis on measures of academic achievement was cited as evidence of continued neglect of the measurement of institutional change. SRI's efforts to develop an acceptable test battery resulted in a delay in the commencement of the 1969-70 evaluation. In fact, testing was not begun until the school year was well underway. In general, SRI's resources were severely taxed by the demands of the evaluation. At the same time, government officials were growing impatient for evidence of the effectiveness of the Follow Through program (Elmore, 1976).

In February, 1970, the results of the first year's evaluation were made available by SRI. While the administrators had expected a tremendous effect from Follow Through, the data did not support these expectations. A meeting of SRI personnel, Follow Through administrators, sponsor representatives, OEO officials, and outside consultants convened to review the first year's results. SRI was severely criticized for its effort. Elmore (1976) detailed three decisions regarding the evaluation that emerged from this meeting. One decision was that the first year in which a model was implemented at any site would be considered an implementation year and data would not be collected for a model at any site during the first year of implementation. The other two decisions are extremely important and reflect the degree to which decisions made by federal officials are controlled by supporters.

It was decided that the evaluation of effects at the end of each year was premature, and that interim assessment would be suspended and analysis would be limited to the entry-exit effects of a child's four year (or, in the case of first grade entry, three year) participation in Follow Through. The effect of this decision was to preclude comparisons between models and to focus the evaluation on the question of overall effectiveness of the Follow Through program. In other words, it was intended to obscure differences between models, when identification of such differences was the purported purpose of the planned variation experiment. The final decision was that the first year report would never be made available to the public. This move can only be construed as a deliberate effort by the government to conceal important information from the public.

## EVALUATION

In 1970, the Office of Education responded to the growing criticism by commissioning an independent review of the evaluation. The conclusion of the review panel was that the evaluation was troubled by a lack of clarification of objectives and purpose for the study. One of the panel members, Marvin Alkin, (1970) claimed that the Office of Education and SRI had failed to specify the kind of study they had intended. Alkin suggested that the most appropriate approach to evaluating the Follow Through planned variation experiment was a summative evaluation in which the relative effectiveness of the various models in achieving the objectives of Follow Through would be determined. Alkin argued that such an approach would provide decision makers with the information needed for the “potential deletion of programs, encouragement of others, or perhaps mandating a particular approach” (p. 5). Alkin’s suggestions went unheeded and the evaluation continued as it had previously.

In October, 1970, SRI met with Follow Through administrators and representatives of the model sponsors to review the results of the 1969-70 school year evaluation. The report, although twice revised, was never considered acceptable and it was never released to the public (Elmore, 1976).

During 1971 Egbert and Snyder both resigned from the Follow Through administration. Rosemary Wilson became the new Director and, in November of that year, Garry McDaniels was appointed the chief of evaluation. According to Elmore (1976) this change in administration resulted in a number of changes in the evaluation. One of the first steps McDaniels took was to contract with an outside consulting firm, the Huron Institute, for technical assistance on the design of the evaluation. The resulting changes brought the evaluation in line with the recommendations Alkin (1970) had previously made; from that point on the evaluation focused on comparison of the effectiveness of the various models.

SRI’s 1969-70 evaluation precipitated what McDaniels described as a “crisis of confidence in the contractor” (quoted in Guttman & Willner, 1976, p. 170). McDaniels elected to put the contract for future data analysis out for competitive bidding. Although SRI continued to collect data, the evaluation was contracted to Abt Associates of Cambridge in June, 1972.

*Abt Associates.* The decision to reduce intermediate testing that was made following the Atlanta meeting in 1968 had been designed to preclude the comparison of models. However, the McDaniels-Huron redesign of the evaluation placed its emphasis directly on the delineation of differential model effectiveness, and made it important to obtain measures at the end of every year of program involvement, not just at entry and exit points. Although it was impossible to obtain intermediate data for children who had entered the program prior to this point, such data collected on children entering Follow Through in 1971 could serve as the basis for a longitudinal evaluation. This group of children became the primary focus of the national evaluation.

Abt Associates conducted annual evaluations and produced yearly reports, which have been published in four volumes titled *Education as Experimentation: A Planned Variation Model* (1977). In each volume, various analytic strategies are used to evaluate different cohort streams. (The term cohort refers to a group of children entering Follow Through in a given year. Cohort I entered in 1969, Cohort II in 1970, Cohort III in 1971. The term stream refers to the grade at which children entered the program. Because

## PROJECT FOLLOW THROUGH

children entered at either first grade or kindergarten there are two streams for each Cohort. Thus Cohort II-EF (enter first) refers to children who began participation in Follow Through in the first grade in 1970; Cohort I-K refers to the group of students who began Follow Through in kindergarten in 1969.)

Volume IV of this report provides the most comprehensive evaluation of the differential effectiveness of the various instructional models. Each sponsor included in the evaluation had at least two years of implementation before Cohort III entered Follow Through. Thus, it may be assumed that to the extent that models were able to be implemented in a number of settings, they would have been by this time. In addition Cohort III was entering Follow Through just as the McDaniels-Huron overhaul of the evaluation began. Consequently, they have been the primary focus of the evaluation since then and analysis of Cohort III data provides the first opportunity for the comparison of model effects. Most commentaries, criticisms, and critiques of the evaluation have pertained to Volume IV of the evaluation report. Therefore, for the purposes of this paper, future mention of the national evaluation will refer to this volume.

The first part of the Abt report, Volume IV-A, *An Evaluation of Follow Through* (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977), includes a summary of the pattern of effects across thirteen sponsored models. The second part, Volume IV-B, *Effects of Follow Through Models* (Boch, Stebbins, & Proper, 1988), devotes an entire chapter to the description of effects for each of seventeen models. For a more detailed discussion of the effects of a given model, the reader is referred to Abt's Volume IV-B. For the purposes of this discussion, effects will be presented in summary form consistent with that of Volume IV-A.

The models were divided into three broad categories according to their areas of primary emphasis. This classification scheme was derived judgmentally and determination of the typology was based on the sponsors' program descriptions and stated goals and objectives of the models (Stebbins et al., 1977). The Basic Skills category included models that focused primarily on directly teaching fundamental skills in reading, arithmetic, spelling, and language. The Cognitive-Conceptual category was composed of models that aimed to develop "learning-to-learn" and problem solving skills. Models in the Affective-Cognitive category emphasized development of self-concept and positive attitudes toward learning, and secondarily on "learning-to-learn" skills. (The Follow Through models have been classified differently by different authors: Maccoby & Zellner, 1970; Gordon, 1972; Emerick, Sorensen, & Stearns, 1973; Parker & Day, 1972; Stebbins et al., 1976).

A brief description of each of the thirteen models included in Volume IV-A of the Abt report follows. More detailed descriptions of the models may be found in Volume IV-B of the report (Bock, Stebbins, & Proper, 1977). The models are grouped according to the Abt typology.

### **Description of Models by Type**

#### ***Basic Skills Models.***

*Direct Instruction*, sponsored by the University of Oregon. The curriculum emphasis of this model was on basic skills in reading, arithmetic, and language. Behavioral methods were used in conjunction with sponsor-developed teaching materials.

## EVALUATION

Lessons were carefully sequenced and specified teacher behavior was scripted. Instruction took place in small, homogenous groups. Children's progress was assessed frequently.

*Behavior Analysis*, sponsored by the University of Kansas. This model emphasized instruction in reading, writing, spelling and math. Teaching took place in the context of a token economy in which tokens were exchangeable for tangible and activity reinforcers. Programmed instructional materials were used. Children's progress was continuously monitored.

*The Language Development (Bilingual Education) Approach*, sponsored by the Southwest Educational Development Laboratory. This model stressed bilingual language development for Spanish speaking children. Positive emphasis on the child's native language and culture was emphasized. Spanish and English were taught simultaneously; teaching procedures were not specified.

*California Process Model*, sponsored by the California State Department of Education. This model stressed the joint home-school determination of educational goals. Individual diagnosis and prescriptive teaching were emphasized. Instructional procedures were not specified.

### ***Cognitive-Conceptual Models.***

*Florida Parent Education*, sponsored by the University of Florida. Curriculum objectives varied depending on the assessed needs of individual children. The emphasis of the model was on training parents as aides who spent half their time as classroom assistants and the other half working in the homes of other Follow Through parents.

*Tucson Early Educational Model (TEEM)*, sponsored by the University of Arizona. The emphasis of this model was development of broad intellectual skills and positive attitudes toward school. Language was emphasized as the medium of skill development. Children's interests determined the curriculum.

*Cognitively Oriented Curriculum*, sponsored by the High/Scope Educational Research Foundation. This model was derived from Piagetian theory. Children scheduled their own activities. Teachers were trained to function as catalysts rather than providers of information. Science, math, and reading were emphasized.

*Individualized Early Learning Program*, sponsored by the University of Pittsburgh. The model emphasized instruction in reading, math, and language. Teaching materials were individually prescribed. Instructional procedures included programmed learning and positive reinforcement.

### ***Affective-Cognitive Models.***

*Mathemagenic Activities Program*, sponsored by the University of Georgia. The objective of this model was to develop initiative and decision making skills. Learning-by-doing and positive reinforcement were stressed.

*Responsive Education*, sponsored by Far West Laboratory. Instruction was self-paced and self-determined. The primary objective of the model was the development of problem solving skills, sensory discrimination, and self-confidence. The model proceeded from the assumption that given self-esteem and an appropriate learning environment, acquisition of academic skills would follow.

## PROJECT FOLLOW THROUGH

*Interdependent Learning*, sponsored by the City University of New York. Primary emphasis was on learning-to-learn. The curriculum was based on small group instructional games which were intended to develop skills in reading, language, and arithmetic.

*Bank Street*, sponsored by the Bank Street College of Education. The curriculum objectives of this model included the development of positive self-image, creativity, coping skills, and the use of language to formulate and express ideas. Instructional procedures were not described.

*Open Education (EDC)*, sponsored by the Education Development Center. The primary objectives of this model were development of self-respect, imagination, openness to change. The schedule was flexible with children initiating and terminating activities. The open classroom approach stressed a stimulating environment. The model assumed that basic academic skills will be more readily acquired if they are not treated as academic exercises.

### Outcome Measures

Ten outcome measures were included the Abt evaluation. These measures were intended to assess performance in different learning domains.

1. *Raven's Coloured Progressive Matrices* (Raven, 1956). This test is generally considered a measure of problem-solving ability in visual and perceptual tasks. The test was shortened and adapted for administration. The test consists of patterns from which one element is missing. The child is required to identify the missing element from an array of available alternatives.
2. *Coopersmith Self-Esteem Inventory* (Coopersmith, 1967). This test is intended to be a measure of the way children feel about themselves, how they think others see them, and their feelings about school. Test items are statements to which the child responds "like me" or "not like me."
3. *Intellectual Achievement Responsibility Scale (IARS)* (Crandall, Katkowsky, & Crandall, 1965). The language of the test was modified for administration. The test is made up of two subtests that are intended to measure the extent to which a child sees self or others as responsible for the child's successes (IARS Positive) or failures (IARS Negative). Test items consist of a description of an event followed by two alternatives from which the child selects a response that describes who was responsible for the event.

The remaining assessment devices were selected from the Metropolitan Achievement Test-Elementary Level, Form F (MAT, 1970).

4. *Word Knowledge*. This test measures how varied the child's vocabulary is. Items are sampled across a broad range of content areas and parts of speech.
5. *Reading*. This is a comprehension test that includes four types of tasks: identifying the topic of a passage, identifying specific information in a passage, determining the meaning of an unfamiliar word from context cues, and drawing inferences from a passage. The inferential items are emphasized in the test.

## EVALUATION

6. *Spelling*. In this test, words are dictated. Test items are selected to include words common to textbooks, and those frequently misspelled.
7. *Language*. There are two parts to this subtest. Part A assesses the identification of punctuation, capitalization, and word usage errors. In Part B the child's ability to recognize declarative, interrogative, and incomplete sentences is assessed.
8. *Math Computation*. This is a test of simple number facts in basic operation. Children must answer test items presented in horizontal as well as vertical notation.
9. *Math Concepts*. This is a test of the child's knowledge of fundamental math principles and relationships.
10. *Math Problem Solving*. This test attempts to assess total math ability. Test items require the use of all four fundamental operations; some multiple-step problems are included. The primary emphasis is on reasoning with numbers and operations. Computational and vocabulary demands are minimized.

Stebbins et al. (1977) grouped the measurement instruments or subtests into three categories in accordance with their interpretation of the sponsor's goals. For example, models emphasizing direct teaching of basic academic skills were expected to demonstrate success on tests that measure these skills. The measures classified as basic skills are the following MAT subtests: Word Knowledge, Spelling, Language (parts A and B), and Math Computation. Models that emphasized more general problem solving skills as instructional objectives were expected to be successful on measures classified as Cognitive-Conceptual. These measures included the Raven's as well as the Reading, Math Concepts, and Math Problem Solving subtests of the MAT. Sponsors that stressed the development of self-concept were expected to have significant impact on measures that were classified as Affective: the IARS (Positive and Negative) and the Coopersmith.

### **Analytic Strategy**

In an effort to compensate for the weaknesses in the quasi-experimental design, Abt Associates subjected the data to a number of different analytical procedures (Stebbins, 1977). The majority of analyses are based on a pre-post test and comparison groups design. As noted earlier, subjects were not randomly assigned to treatment and comparison groups. Instead, SRI had constructed comparison groups that were as similar as possible to the Follow Through sample. However, earlier analyses (e.g., Emerick, Sorensen, & Stearns, 1973; Cline et al., 1974, 1975) suggested that the degree to which the selected comparison groups were in fact similar varied from sponsor to sponsor and site to site. This noncomparability of groups was the primary concern of Abt in selecting strategies for evaluation of the data (Stebbins et al., 1977). In an effort to compensate for this design weakness, the comparison of Follow Through and non-Follow Through students was approached in a variety of ways. The two primary analyses were, (a) comparison of students at a given Follow Through site with a local comparison group (selected non-Follow Through students within each site), and (b) comparison of Follow Through students at a given site with a nationally pooled comparison group (non-Follow Through groups pooled across all sponsors and sites). In the pooled analysis the same set

## PROJECT FOLLOW THROUGH

of children serve as the comparison group for each sponsor. For both local and pooled comparisons statistical adjustment of obtained measures were made based upon relevant background variables.

### Summary of Effects

For each outcome subtest, Abt evaluators compared the performance of Follow Through children in a given site group with designated comparison groups. This process resulted in more than two thousand comparisons (Anderson, St. Pierre, Proper, & Stebbins, 1978). The results of the analysis are complex and consume several volumes totaling in excess of 2,000 pages. A complete description of effects for each model is presented in Volume IV-B (Bock, Stebbins, & Proper, 1977). The effects were summarized and presented in Volume IV-A (Stebbins et al., 1977). The results reported here are based primarily on the patterns of effects described in that volume. The reader desiring more specific information and detail is referred to the report in its entirety.

The patterns of effects are based on data collected on children who entered Follow Through in 1970 (Cohort II) and 1971 (Cohort III). Two streams of children are included in each cohort, those entering in the first grade (EF) and those entering in kindergarten (K). Thus, the data base is comprised of information on a total of four streams of children.

The basic tactic of summarizing data was to average the scores of individual children on each of the outcome measures to yield a "group" score. The term group refers to all the Follow Through children in a model's local project in one of the four cohort streams. For example, Responsive Education's cohort III-K children in Deluth constitute a group. These group outcomes then served as the unit for calculating all effect averages.

For each outcome variable, the adjusted group score was compared to both the local comparison group and the pooled comparison group. The difference between treatment and comparison groups was used as the measure of effect. An effect was judged to be materially, or educationally, significant if it met the following two-step criterion: 1) the difference in adjusted scores was statistically significant at the .05 level, and 2) the difference was at least one-fourth standard deviation (approximately 2 months difference at the third grade level). Thus, when Follow Through adjusted scores on a given outcome variable exceeded non-Follow Through adjusted scores by at least .25 standard deviation *and* when the difference was statistically significant, the outcome was considered positive. Conversely, when non-Follow Through adjusted scores surpassed Follow Through scores by the same criteria, the outcome was considered negative. When the differences between scores on a given variable failed to meet the established criterion, a null effect was declared.

Abt contended that, despite the method of selection, the Follow Through and comparison sites were generally well matched. In the majority of cases, differences between Follow Through and non-Follow Through sites were distributed in a manner consistent with the assumptions of the analysis of covariance. However, in some instances, the differences between treatment and comparison sites were more pronounced. The use of analysis of covariance in cases where the initial mismatch is substantial has been a source of controversy among statisticians. While the Abt evaluators declared confidence in the covariate set, they opted to eliminate effects based

## EVALUATION

on extreme values of adjustment in order to avoid the possibility of analytical error or statistical artifact (Stebbins et al., 1977). Thus, all effects in which the covariance adjustment exceeded one-half of a standard deviation of the outcome variable were omitted from the summary. The application of this exclusion rule has been criticized (e.g., Becker & Carnine, 1981) as resulting in the attenuation of differences between models.

Average model effects were computed in the following manner. The total number of negative outcomes for a particular category of measures (e.g., basic skills) was subtracted from the total number of positive outcomes and the difference was divided by the total number of comparisons (local and pooled). The average effects of a type of model (e.g., basic skills) were similarly calculated. Averages for individual models as well as those for model type combine the local and pooled estimates of effects across all four cohort streams.

Abt Associates identified five major findings related to the question of differential model effectiveness. These findings are paraphrased from Volume IV-A (Stebbins et al., 1977, pp. 135-148).

*Finding 1: The effectiveness of each Follow Through model varied substantially from site group to site group; in comparison overall model averages varied little.*

The range of effects between sites within each model was greater than the range of average effects between models. Every model had at least one group with a negative average effect in each outcome domain relative to the comparison groups. In addition, every model had at least one group that demonstrated a positive average effect compared to non-treatment groups in at least one outcome domain. It is curious that, despite the fact that it provides no information whatsoever about the differential effectiveness of various educational methods, this observation has been heralded as the most important finding of the Follow Through evaluation. Stebbins et al. (1977) concluded that the finding provides evidence that local circumstances are more clearly related to academic achievement than are instructional methods (e.g., Anderson, St. Pierre, Proper, & Stebbins, 1978; House, Glass, McLean, & Walker, 1978). There is, however, no empirical evidence in the Abt evaluation of a functional relationship between any "unique local conditions" and outcome measures. Thus, it must be concluded that this statement is pure conjecture (Bereiter & Kurland, 1981).

The Abt evaluators suggest that there is evidence that every model can be effective. However, there is not much benefit to be gained by an instructional approach that is so sensitive to local variations that it can be successful only under severely restricted conditions. The purpose of implementing the models in a range of communities was to determine what models were most effective. It may be assumed that effectiveness is related to generalizability. Thus, to the extent that a model was effective in a variety of settings, it may be judged to be of more practical value.

Furthermore, to the extent that there are differences between sites, these differences may be considered to be distributed among the models in a non-systematic fashion. That is, local circumstances affect the degree to which a model is implemented. No one would deny that is the case. Further, to the degree that implementation varies,



## PROJECT FOLLOW THROUGH

results may be expected to vary. The possibility must be entertained then that variability from site to site was a function of variables affecting implementation.

Other researchers have subjected the data to alternative analyses and offer a different perspective on site variability. For example, Bereiter and Kurland (1981) sought to determine what proportion of the variance not accounted for by entering characteristics could be attributed to model differences. After parceling out variation due to covariates, they found that model differences explained up to 55 per cent of the variance, depending on the particular outcome measure considered and on the covariate set. In the most extreme case, the difference between highest and lowest model was 3.6 standard deviations. In all cases, the difference was at least 1.4 standard deviations. In short, these data dispute the claim that variance due to treatment was insignificant.

*Finding 2: Models that emphasize basic skills succeeded better than other models in helping children gain these skills.*

The average rank of models in the “basic skills” category was significantly higher than the other two model types on measures of basic academic skills. In other words, groups of children in Basic Skills models performed significantly better on measures of academic skills than did non-Follow Through groups. Abt evaluators concluded that a Basic Skills model would be preferable if an educator were concerned with teaching skills such as spelling, math computation, language, and word knowledge.

As is the case any time data are aggregated, a certain amount of information regarding the effectiveness of individual models is obscured by the summarization procedure. For example, in the basic skills domain neither the California Process model nor the Language Development model had positive average effects. Yet when the site group effects of these models are combined with those of the Direct Instruction and Behavior Analysis models (both of which had positive average effects) the overall average effect for this category is superior. In other words, the method of summarization is advantageous for some models.

While the validity of the Abt taxonomy has not been contradicted by more formal methods of category construction, such as factor analyses (Camilli, 1980), it appears that categorical presentation of results does not provide the most accurate description of the results. This is particularly true when the basis for inclusion in a given category is not specified. It is not inclusion in a category that leads to educational effectiveness, but rather, the particular instructional procedures used. Thus in the absence of definition of the commonality of variables necessary for class inclusion, more meaningful information is provided by looking at the effects of a given instructional method.

On measures of basic skills, only the Direct Instruction and Behavior Analysis models had positive average effects. The Direct Instruction model had an unequivocally higher average effect on scores in the basic skills domain than did any other model.

*Finding 3: Where models have put their primary emphasis elsewhere than on the basic skills, the children they served have tended to score lower on tests of these skills than they would have done without Follow Through.*

All models other than those labeled “Basic Skills” had more negative than positive outcomes on measures in the basic skill domain. Intervention with any other type

## EVALUATION

of model resulted in an iatrogenic effect. This means that when instruction focused on objectives other than the acquisition of basic skills, the performance of students in the *comparison* groups was *superior* to that of Follow Through students. The excess of negative over positive scores on basic skill measures was 19% for Affective models and 17.2% for the Cognitive/Conceptual models. These data suggest that disadvantaged children are more likely to acquire basic academic skills by the end of third grade if they received instruction typical of traditional elementary classrooms than they would if they were subjected to the majority of educational methods evaluated in Follow Through. This finding is especially important in light of the fact that it was precisely because disadvantaged students have consistently failed to benefit from traditional educational practices that they were singled out for participation in Follow Through. Thus, the four years these students spent in Follow Through actually resulted in an exacerbation of the educational deficit it was intended to remediate.

Because Abt's interpretation implies that the majority of Follow Through models were actually detrimental to disadvantaged children, this finding deserves to be carefully scrutinized. Some critics of the Abt evaluation (e.g., Camilli, 1980; House et al., 1978) have argued that failure to randomly assign students to treatment and comparison groups leads to systematic bias of outcome measures when the covariate set is imperfect. In a case in which the treatment group is initially disadvantaged relative to the comparison group, the expected result is under-adjustment. In such a situation, the outcome will appear to favor the comparison group. Critical to this argument is the assumption that there are systematic differences favoring the comparison group. Abt responded by describing the entering characteristics of the groups as follows:

In 47% of the sites included in the evaluation, the treatment group and comparison group pretest means were within one quarter standard deviation of each other. In 77% of the sites, they were within one half standard deviation. More importantly, *the differences were not particularly biased toward either group*. Treatment group means were above comparison group means in 44% of the sites, while the opposite was true in 56% of the sites [emphasis added]. (Anderson et al., 1978, p. 166)

Based on Abt's description of the differences between groups, it may be concluded that the groups, as a whole, were not grossly mismatched. Furthermore, an explanation of the negative results as regression artifacts would not apply to cases where the measurement of covariates is error-free (Campbell and Baruch, 1975). The Abt evaluators assume "perfect reliability . . . for all of the covariates in this study" (Stebbins et al., 1977, p. 93). There is no evidence that the assumptions of the analysis of covariance have been violated. Consequently, there would appear to be little basis for assuming that the negative effects were artifacts. Abt's conclusion was that, with the exception of Basic Skills models, Follow Through students' achievement scores at the completion of the third grade were in fact lower than they would have been in the absence of this "compensatory" education.

*Finding 4: No type of model was notably more successful than the others in raising scores on cognitive conceptual skills.*

## PROJECT FOLLOW THROUGH

No model type had an overall average positive effect on measures in this domain which includes reading comprehension and problem solving. Only three models seem to have had any considerable impact on these complex skill areas. Two of these models are Basic Skills models (Direct Instruction and Language Development (Bilingual) Approach) and one is an Affective-Cognitive model (Mathemagenic Activities). No model in the Cognitive-Conceptual category obtained a positive average effect on these measures despite the fact that their instructional programs emphasized development of these types of skills. In other words, if the Follow Through models that focused on more advanced cognitive conceptual skills are capable of influencing standardized measures of those skills, they failed to do so after four years of instruction.

*Finding 5: Models that emphasize basic skills produced better results on tests of self-concept than did other models.*

On the average, children in Basic Skills models performed better on affective measures than did children in Cognitive-Conceptual or Affective Cognitive models. All four of the Basic Skills models had positive average model effects on affective measures. The only other model to demonstrate a positive average effect was the Florida Parent Education model. In every case, the models that focused on affective development had negative average model effects of measures in this domain.

The Direct Instruction and Behavior Analysis models ranked first and second, respectively, in average effects on affective measures. Both of these approaches stress careful structuring and sequencing of curriculum materials that are designed to limit the number of errors and to ensure successful performance. In addition, they both rely on frequent measurement of the child's progress in order to provide immediate remediation. These models view positive self-concept as a consequence of acquisition of basic skills. In contrast to the proposition that self-concept is a necessary antecedent condition for learning, they hold that instruction that results in academic success will lead to improved self-concept. The data appear to support this theory.

However, again, it is misleading to make the claim that instruction in a Basic Skills model leads to academic success and improved self-concept. Significant differences on both categories of measures were observed for only two of the four Basic Skills models, Direct Instruction and Behavior Analysis. The other Basic Skills models did have positive average effects on measures in the affective domain but they had negative average effects on measures of basic skills. In addition, the Parent Education model was the only other model to demonstrate a positive effect in both basic skills and affective development. Thus, describing the result as a "Basic Skills" effect does not provide useful information about the specific instructional variables that give rise to significantly better performance in both outcome areas. If the models had been more precisely defined, a component analysis might make it possible to delineate the factors common among models that obtained similar results. In the absence of such information, it is not especially useful to attribute the observed effects to a categorical label. Significant differences on both categories of measures were observed for only two of the four Basic Skills models, Direct Instruction and Behavior Analysis. The other Basic Skills models did have positive average effects on measures in the affective domain but they had negative average effects on measures of basic skills. In addition, the Parent

## EVALUATION

Education model was the only other model to demonstrate a positive effect in both basic skills and affective development.

### Alternative Analyses

#### The House Critique

That the evaluation of Follow Through would be contested was to be expected given not only the main findings themselves but the overall context of the experiment. Preliminary reports by Abt Associates (Cline et al., 1974; Cline et al., 1975; Stebbins, et al., 1976) suggested that the effects of the model sponsors were not homogenized but that different sponsors yielded different performances. In particular, it appeared that models whose instructional emphasis was on cognitive-conceptual or affective development were less effective, as measured by the selected outcome variables. These findings prompted a privately funded review of the final evaluation. In 1976, even prior to the publication of Abt's final report (Stebbins et al., 1977), the Ford Foundation awarded a grant to the Center for Instructional Research and Curriculum Evaluation at the University of Illinois. Dr. Ernest R. House was named director of the project, whose task was to examine the analyses and conclusions of the Abt report and to provide an independent assessment and interpretation of the data. Because the report published by House and his staff (House et al., 1978) constitutes the only serious challenge to the findings of the national evaluation, its contents will be considered in some detail.

House et al. (1978) found problems in three general areas: classification, measurement, and analysis. While all aspects of the critique will not be commented upon here, there were some challenges made about the validity of the Abt findings that must be addressed in order to present a clearer picture of what was actually learned about the education of disadvantaged young children.

The House report devoted considerable space to a discussion of their concern about the Abt typology of models and outcome measures. Recall that, for the purposes of summarizing a mass of data, Abt Associates had categorized the subtests of the outcome measures as Affective, Cognitive-Conceptual and Basic Skills depending on the particular types of skills each subset was presumed to measure. They also categorized the models according to their judgment of each models primary instructional emphasis.

The subtests of the Metropolitan Achievement Test fell into two categories. Word Knowledge, Word Analysis, Spelling, and Mathematical Computation were referred to as "basic skills;" while Reading, Mathematics Concepts, and Mathematics Problem Solving were labeled "cognitive-conceptual skills". This division was made by the national evaluators based on the notion that basic skills could be taught directly while cognitive-conceptual skills may or may not be taught directly but may be acquired indirectly as a result of various experiences (Stebbins, et al., 1977, p. 37). House and his associates argued that the distinction may be based on the following: the subtests classified as cognitive-conceptual "require something other than simple memory" (House et al., 1978, p. 137). House extends this argument to say that such a classification is more favorable to models that emphasize "rote learning" and that it implies that skills taught by rote methods are basic and therefore of greater importance than other skills. In order to avoid what they consider undue delegation of importance to skills such as spelling, math

## PROJECT FOLLOW THROUGH

computation, and word analysis and knowledge, the House group proposed that these skills be labeled as “mechanics of reading writing, and arithmetic, noting that mechanics can be taught by rote methods.” (House et al., 1978, p. 137).

The argument here appears to be that the models that made the greatest impact on disadvantaged children’s ability to decode and recognize words, compute arithmetic problems, and spell words did so through rote methods. This statement assumes that there were models that used rote methods; presumably House was alluding to the Behavior Analysis and Direct Instruction models (Bereiter & Kurland, 1981). Neither of these models, however, have anywhere described their teaching method as rote. In fact, the Direct Instruction model had explicitly denounced rote memorization (Bock, Stebbins, & Proper, 1977, p. 65). Both the Behavior Analysis and Direct Instruction models use teaching methods that have been derived from years of experimentation in the analysis of behavior. Greer (1982) has discussed the naiveté of the majority of educational researchers about the advances this research tradition has made toward a technology of teaching. The fallacy committed by House et al. (equating the methods used by these two models with rote memorization) is not uncommon.

A more serious implication is that the MAT measures are biased in favor of these models that have mistakenly been said to use rote methods. In order for the selection of outcome measures to be considered biased it would need to be shown that outcome measures were used that assessed skills that only some subset of sponsors thought were important. House et al. (1978) proclaimed that both the public and professionals would agree that all skills measured by the MAT are basic to mathematical skill and literacy. In addition, virtually all of the models taught reading, writing, spelling, and arithmetic; the sponsors, by their acceptance of the assessment battery, tacitly agreed to the appropriateness of these outcomes. The MAT then may be assumed to measure skills that were universally agreed upon as important for children to acquire by the third grade and its use cannot be assumed to be biased toward any model. Bereiter and Kurland (1981) pointed out that the only other way in which the use of the MAT could constitute bias was if the test were constructed in such a way that it gives preferential treatment to children who learned the skills by one method of instruction (e.g., rote) over some other process. Since it is unlikely that House et al. assumed this to be the case, the charge that the MAT favors rote methods is unfounded.

A second criticism related to the selection of outcome measures is that the national evaluation was “unfair” because outcome measures did not measure all of the goals of all the sponsors. The evaluators readily admit that “no common battery could be developed that would encompass all the various sponsors’ goals and objectives” (Stebbins et al., 1977, p. 10). Anderson et al. (1978) addressed the impracticality of the House committee’s criterion for fair evaluation when they declared:

Any program that wishes to rid itself forever of the discomforts of evaluation need only add to its list of objectives one metaphysical, obscure, or otherwise unmeasurable purpose (say, the improvement of “life chances,” a phrase of great importance in compensatory education legislation). Since such an outcome can be assessed only obliquely and imprecisely, if at all, any evaluation of that program must thenceforth be an “unfair evaluation.” (p. 163)

## EVALUATION

The Abt evaluators acknowledged the fact that basic skills were more adequately sampled by the selected outcome measures: "Whereas the basic skills measures are probably a reasonable battery for examining achievement which might be expected in that learning domain, the measures on the cognitive and affective domains are much less appropriate" (Stebbins et al., 1977, p. 35). Because of these limitations in the test battery, House et al. (1978) urged readers of the report to "assume that the national evaluation amounts essentially to a comparative study of the effects of Follow Through models on the mechanics of reading, writing, and arithmetic" (p. 145). Even if followed, this perspective still provides invaluable information from the experiment as will be seen in the following section when the results of an alternative approach to evaluating model performance by Bereiter and Kurland (1981) are considered.

The most serious challenge House et al. (1978) leveled at the national evaluation was that the reported patterns of effects did not really exist; that there were no differential model effects on outcome measures. In an attempt to prove this point, House and his colleagues conducted two forms of reanalysis of the data.

The first was a weighted aggregation of the numerical mean site effects (Anderson et al., 1978). Abt analysts had used the child as the unit of analysis. Because the number of children enrolled in Follow Through varied from site to site, and because the number of children at a site will effect statistical significance, House et al. argued that the dual criteria used by Abt to define effect potentially confounded model effect with the number of students in a given site. The House committee computed the effect at each site in the following manner. Using the data provided in the appendices of the Abt report (*Education as Experimentation*, 1977) they calculated the average difference between Follow Through and non-Follow Through adjusted means from the local analysis only and divided the difference by the standard deviation of the outcome variable. They then rank ordered the models based on their average effects on the MAT subtests, disregarding statistical significance, and compared the obtained order with that of Abt.

The results were largely comparable with the national evaluation (House et al., 1978, p. 150). (According to Anderson et al., 1978, Spearman's rho equals .78 between the two sets of ranks.) In both cases, the instructional model that had the greatest average effect on MAT measures was the Direct Instruction model. The average effect for Direct Instruction averaged across total reading, total math, spelling, and language effects was .272; that of the second ranked model was .038. While the Direct Instruction model was ranked first in both cases, there were some changes in the relative standing of some other models between the two analyses. The differences in rank that did occur may be attributed to differences in analytic decisions (Anderson et al., 1978; Wisler, Burns, & Iwamoto, 1978). First, the analysis conducted by the House group used only data from Abt's local analysis of covariance; Abt used both the local and the parallel pooled analysis to calculate average effects. Second, the House reanalysis included data from only two of the four cohort streams. Third, only four MAT subtests were included in the averages of the House group; Abt used all eight subtests in their averages (both basic skills and cognitive-conceptual) as well as the Raven's. Fourth, Abt had omitted effects when the size of adjustment was excessive; House et al. included these questionable effects in their analysis. Finally, the method of computing effects differed between the two studies.

## *PROJECT FOLLOW THROUGH*

The second form of reanalysis to which House et al. subjected the data was a site level analysis of variance, which they contrasted with the Abt analysis that used individual children as observations. Replacing the child as the unit of analysis with the site had the effect of greatly decreasing the number of cases. The sample sizes in a site analysis ranged from two to eleven, and not surprisingly, the results of this analysis indicated no significant differences between models. The Abt evaluators had considered doing a site level analysis, but rejected it because the inadequate sample size could not support such an analysis (Stebbins et al., 1977) and the null effect would be predetermined. The House group claim that their analysis indicates that “none of the findings accepted by the Abt Associates evaluators as statistically significant holds up as significant when reanalyzed with a different unit” (House et al., 1978, p. 177). Their reanalysis did not however, prove that no differences existed between models. At best, it demonstrated that the Abt evaluators were correct in assuming that a site level analysis of variance lacked the power to reveal significant differences between models.

In general, the findings of the House report do not discredit, but rather confirm the findings of the Abt evaluation. The fact that the qualitative findings of the House report do not differ substantially from findings of the Abt evaluators does more to support than deny the patterns of effects identified in the national evaluation. The bulk of the House critique amounts mostly to a warning to the reader that one should be cautious in interpreting the Follow Through data. Given the complexity of the evaluation, this caution is warranted. It certainly was made explicit by the Abt evaluators, in fact, the majority of issues raised by House et al. were addressed in the Abt report. House et al., however, were correct in promoting various analyses of the data. Different analytical methods may in fact yield contradictory findings. In a situation such as Follow Through, where the design may be best described as quasi-experimental, it is unlikely that any one analysis will present the complete story. A reasonable strategy for discerning the pattern of effects would be to consider a number of alternate analyses and draw conclusions based on the combined findings. Two other major reanalyses of the Follow Through data will be reviewed.

### **Bereiter and Kurland**

Bereiter and Kurland’s (1981) approach to the evaluation of Follow Through differed from that of both Abt and the House committee. Two most frequently cited shortcomings of the Follow Through evaluation are the adequacy of the outcome measures and the noncomparability of control groups. Although the technical adequacy and appropriateness of the affective measures has been questioned, there is considerable agreement (e.g., Stebbins et al., 1977; House et al., 1978) that the MAT was an appropriate measure of achievement. Bereiter and Kurland therefore limited the questions asked of the data to those related to academic achievement as measured by MAT subtests.

Furthermore, Bereiter and Kurland elected to avoid the problem of noncomparable comparison groups and the complexity that results from adjusting for those differences statistically. In their creative and elegant reanalysis of the Follow Through data, Bereiter and Kurland elected to compare the Follow Through models

## EVALUATION

directly with one another to answer questions regarding how the models compared on the various achievement test scores at the end of the third grade.

Bereiter and Kurland (1981) agreed with the House committee that the site was the appropriate level of analysis. However, they identified several methodological weaknesses that resulted in maximum error variance (p. 5). First, the analysis compared the Follow Through models with one another based on differences between treatment and comparison groups. Second, the differences were based on the local analysis, the weakest of the Abt analyses due to mismatch between Follow Through groups and local comparison groups. Third, the analysis was based on data for only two cohort streams.

Bereiter and Kurland conducted a site level analysis that differed from that of House et al. in the following ways (pp. 5-6):

- 1) Site means for Follow Through groups were used as the dependent measure with other site level scores as covariates. (House et al. used site-level differences between Follow Through and non-Follow Through groups as the dependent variable, with individual scores as covariates.)
- 2) When non-Follow Through scores were used, they were used as covariates.
- 3) All data from one site were combined as a single observation, regardless of cohort.
- 4) The analysis was restricted to models having data on 6 or more sites in order to retain the power of the statistical test.

The dependent variable was the mean score for a site on subtests on the MAT averaged over all students in Cohorts II and III for whom data were available. The models were compared by analysis of covariance.

Bereiter and Kurland (1981, pp. 6-13) reported the results for two analyses of covariance: a "full" analysis of covariance, and a "conservative" analysis of covariance that eliminated certain covariates for empirical or rational reasons. Other analyses that varied the covariates were also done; the adjusted scores for these analyses consistently fell between those for the "full" and "conservative" analyses, indicating that the two analyses reported represent the entire range of findings. The difference between models was statistically significant at the .05 level or beyond on every achievement measure, regardless of the analysis. Bereiter and Kurland reported that the performance of any model tended to be consistent across achievement variables suggesting that there is no basis for concluding that certain models are better at one thing while others are better in other areas.

Their main finding was that while some models fluctuated in relative standing depending on the choice of covariates, there was a consistent pattern to emerge from the data. Two models, Direct Instruction and Behavior Analysis, were at or near the top on every achievement variable regardless of the covariates used. Two models, EDC Open Education and Responsive Education, were at or near the bottom on every achievement variable regardless of the covariates used. In the majority of cases the differences between the top two and bottom two models were statistically significant.

Bereiter and Kurland's results also provide some insight into the question of variability between sites. Depending on the achievement variable considered and the particular covariates used, model differences were found to account for between



## PROJECT FOLLOW THROUGH

approximately 17 and 55 percent of the variance between sites not attributable to entering characteristics controlled for by covariates. Bereiter and Kurland present the most extreme case as an example (p. 11). The adjusted mean score of Direct Instruction sites on the Language Part B subtest of the MAT is 3.6 standard deviations above the adjusted mean score of the EDC Open Education sites on the same measure. These data suggest that the between-site residual variability is enormous compared to the differences between sites within a given model. Bereiter and Kurland (1981) disagree with the conclusions of Abt and the House group and declare that they did

... not find variability among sites to be so great that it overshadows variability among models. It appears that a large part of the variability observed by Abt and House et al. was due to demographic factors and experimental error. Once this variability is brought under control, it becomes evident that differences between models are quite large in relation to the unexplained variability within models. (p. 13)

Gersten (1984) used a variety of methodological approaches to examine the variability issue and obtained results that generally supported the conclusions of Bereiter and Kurland. The combined findings of these alternative analyses strongly suggest that site variability can be explained and understood, and that the differences between models are real and related to particular types of instructional procedures.

The analytic approach of Bereiter and Kurland permits an interpretation of results that does not depend in any way on the typology of models or measures that have been objected to (e.g., House et al., 1978), nor have individual model effects been obscured by aggregation. Bereiter and Kurland's analysis revealed clear patterns of differential effectiveness between models. The findings, with respect to achievement, are not inconsistent with those of Abt or House et al. The two models that Bereiter and Kurland identified as having the least impact on achievement scores relative to the other models were EDC Open Education and Responsive Education. These models ranked fifth and twelfth, respectively, on the Abt analysis. When ranking was recalculated by House et al. the Open Education model was tenth in relative standing and Responsive Education ranked seventh. Both of these models were categorized as Affective-Cognitive models by Abt. This model type had the lowest average effect on measures of basic skills according to Abt's analysis.

The two models Bereiter and Kurland found to demonstrate a superior performance on achievement variables were Direct Instruction and Behavior Analysis. The Direct Instruction model was first in relative standing on both the Abt and House ranking. The Behavior Analysis model was ranked seventh by Abt and third by House et al. on achievement measures. Both of these models were considered "Basic Skills" models according to the Abt typology. Models in this category had the highest mean effects on measures of achievement.

### **Kennedy**

Kennedy (1978) obtained findings that substantiate those of Bereiter and Kurland in her reanalysis of the Follow Through data. Kennedy aggregated the data as follows. Each estimate of each site effect was standardized using the pooled within-project unadjusted standard deviation of children's scores. Kennedy determined the average of

## EVALUATION

all effects estimated for each site (both pooled and local analyses), using sites in Cohorts II and III and in both entering streams (K and EF). Effects considered to be untrustworthy (using Abt's criterion) were eliminated. The outcomes reviewed in Kennedy's report are limited to nine major models and to five outcome measures; the Language, Reading Comprehension, Math Computations, and Math Problem Solving subtests of the MAT, and the Raven's Progressive Matrices Test. (Kennedy restricted the number of outcomes to achieve parsimony. She noted that her results did not point to conclusions that were greatly discrepant from those of the Abt evaluation which aggregated results across groups of outcomes.)

The results indicated that four models (Responsive Education, Tucson Early Education, Florida Parent Education, and the Language Development (Bilingual) Program) showed no sizable effects on any of the five outcome measures. In other words, there were no important differences between Follow Through and non-Follow Through students on MAT measures or the Raven's after four years of participation in a Follow Through site.

Three models (Bank Street, Cognitive Curriculum, and EDC Open Education) also had a preponderance of null outcomes. In addition, however, these models had at least one average negative effect greater than  $-.25$  SD. The findings for these seven models support Abt's conclusion that for most models, the children they served scored about the same as disadvantaged children would be expected to score without Follow Through. On some outcome measures, the performance of Follow Through students was actually below that which would be predicted without intervention.

In contrast, two models demonstrated at least one substantial average positive effect. The Direct Instruction model had an especially strong showing on the Language measure and also had a sizable effect on the Math Computation and Math Problem solving subtests. The Behavior Analysis model had its biggest impact on the Math Computation measure, and also had a smaller positive effect on Language. This is consistent with Bereiter and Kurland's (1981) finding that the Direct Instruction and Behavior Analysis models were most effective on achievement measures.

Kennedy made the observation that the models with sizable negative effects were all relatively unstructured models. On the other hand, the two models that had significant positive average effects on some outcome measures were both structured approaches to instruction. This pattern corresponds with the findings of other research pertaining to Follow Through (e.g., Stallings & Kaskowitz, 1974; Bereiter & Kurland, 1981) as well as with the findings of independent studies (Wright, 1975).

### Overall Results

Abt Associates acknowledged that in evaluating research that deviates from a true experimental design, any single analytic treatment of the data is subject to criticism. This criticism is likely to increase in intensity as the "stakes" increase. The implications of Follow Through for the educational community were great. It is not, therefore, surprising that the findings of the national evaluation have met with repercussions. However subsequent reanalyses employing different analytic tactics may assist in determining the degree to which the original findings may be considered valid (Stebbins et al., 1977). The Abt evaluators anticipated some of the alternative approaches to

## PROJECT FOLLOW THROUGH

evaluation that may be useful in clarifying treatment effects. Their multiple strategies analysis has approached the data in a number ways. In addition, they made the data available that would permit others to conduct their own analyses of the data. When the results of different analyses are in agreement, then the confidence placed in the findings may be greater.

The national evaluation of Follow Through, as well as three major reanalyses of the Follow Through data have been reviewed here. With respect to achievement measures, there has not been significant disagreement among these analyses. It should be clear that despite varying approaches to the data analysis, use of different subsets of the analytic sample, and different methods of aggregation, all results converge on finding that *“The highest mean scores on the MAT subtests were attained by students enrolled in two models, the Direct Instruction Model and the Behavior Analysis Model”* [emphasis in original] (Rhine, 1981, p. 302). The Follow Through experiment was intended to answer the question “what works” in educating disadvantaged children. If education is defined as the acquisition of academic skills, the results of the Follow Through experiment provide an unequivocal answer to the question. The evidence provided by the Follow Through experiment clearly indicates that the instructional methods employed in the Direct Instruction and Behavior Analysis models are most effective in teaching the skills necessary for basic literacy and mathematical competence.

*USING THE RESULTS*

## PART IV: USING THE RESULTS

### Dissemination

The purpose of the Follow Through planned variation experiment was to identify effective educational methods. The final criterion of effectiveness must be adoption, as it does the public little good to “identify” little-used methods. No method, effective or otherwise, can be adopted if it is not made accessible to the public. In spite of this rather obvious point, however, the initial planning of Project Follow Through did not provide for any mechanism for disseminating models that were demonstrated to be effective.

The need to make information about effective compensatory education programs available to school districts was recognized by the Office of Education. In 1972 John Evans, head of OE’s Office of Planning, Budgeting, and Evaluation (OPBE) announced a plan that would use all of OE’s resources to identify and disseminate effective educational programs. The plan resulted in the creation of the Joint Dissemination Review Panel (JDRP) and the National Diffusion Network (NDN). While the JDRP and NDN are the primary means of disseminating information about the Follow Through models, dissemination is not restricted to Follow Through models nor is it limited to those models that were most effective in the Follow Through evaluation. Thus, although the stated purpose of these organizations is to disseminate effective technology, in actual practice methods may be disseminated that are ineffective in increasing academic achievement.

This curious situation arose in part from the criteria adopted by the JDRP as the standard of effectiveness. These criteria included “Is the evidence believable and interpretable?” as well as the usual criteria of positive change and consistent effects. According to the JDRP, the positive impact of a program need not be related directly to academic achievement but may include improvement in areas such as self-concept, attitude, and mental or physical health. In addition, a program may be judged effective if it has a positive impact on individuals other than students, for example if it results in improved instructional behavior of teachers.

In 1977 Follow Through sponsors submitted programs to the JDRP. (Note that local projects are submitted, not models. The Behavior Analysis model, for example, submitted eight of their twelve programs for review. All of these were validated as “exemplary and effective.”) Of all those programs submitted by Follow Through sponsors, 22 were validated as exemplary. However, among those “exemplary programs” were programs that had been incapable of demonstrating improved academic achievement in the Follow Through evaluation.

Once a program was validated as effective, it was packaged and disseminated to school districts through the National Diffusion Network (NDN). The purpose of this organization is to help educators implement instructional programs in their schools by providing information about the programs, as well as materials and technical assistance. NDN dissemination is not limited to Follow Through programs and their annual publication includes listings for programs judged (by JDRP) exemplary in twelve areas ranging from vocational education to programs for the gifted and talented. Validated Follow Through programs were established as Follow Through Resource Centers,

## USING THE RESULTS

supported by the Office of Education. These centers were designed to provide information about the model program, to receive visitors, and to conduct training in their respective method.

The JDRP's validation practices did not go unchallenged. The former Commissioner of Education, Ernest Boyer, once wrote to Senator Packwood, "Since only one of the sponsors (Direct Instruction) was found to produce positive results more consistently than any of the others, it would be inappropriate and irresponsible to disseminate information on all the models . . ." (quoted in Carnine, 1984a, p. 87). However, Commissioner Boyer's concerns were ineffective in preventing the widespread dissemination of instructional approaches that were incapable of raising the achievement scores of Follow Through children. This appears to be because the JDRP felt constrained to be "fair" and to represent the multiplicity of goals of education. It should be clear that these practices make it impossible for school districts to discriminate effective from ineffective programs and seem to defeat the very purpose for which the JDRP and NDN were established.

### Funding Decisions

The effectiveness of Project Follow Through may also be measured by the extent to which the findings have influenced decisions regarding funding of Follow Through sponsors. While all Follow Through models have received budget cuts over the years, the disbursement of available funds was based not on any criteria related to effectiveness of the program but on a non-competitive continuation basis (Bell, 1983). In fiscal year 1982, the funding formula was changed so that sponsors with JDRP-validated programs received the lowest level of funding while the highest level of funding went to those sponsors who had not been validated (Ramp, 1983).

Apparently, the rationale behind this practice was that programs that had already proven themselves did not require as much money as did non-validated programs. More importantly, the decision suggests that the remaining programs *would* reach the criteria for validation given more money. This notion was expressed in a letter from then Acting Assistant Secretary for Elementary and Secondary Education, Dr. Jean Benish, dated July 6, 1982, which stated in part, "the amount of a grant (to a sponsor) based on an unvalidated project was slightly higher than the amount based on a validated project. One of the priorities for the program during the next school year is to validate as many additional qualified projects as possible." (quoted in Carnine, 1984a, p. 88).

The logic that funding ineffective programs at a higher level would make them effective was faulty. Not one additional program was validated during the following school year. Yet the same funding policy was implemented the following year, again differentially funding ineffective programs at a higher level. To this day, these programs have never been validated and it is not likely that they ever will be. The data do not suggest that increased financial support of these programs will lead to increased academic performance in their students. Instructional practices, not dollars, appear to be critically important.

The results of the Follow Through evaluation also failed to influence decisions regarding disbursement of federal research dollars. For example, Carnine (1983) submitted a document to the House subcommittee hearing on Follow Through

## *PROJECT FOLLOW THROUGH*

Amendments of 1983 that asserted that the government had made decisions that discriminated against effective Follow Through programs. This accusation was directed at the Administration for Children, Youth, and Families (ACYF), a federally funded agency designed to work with local elementary schools and Head Start projects to improve basic educational skills. The funding policy for ACYF's Basic Skills Program denied an opportunity for behaviorally oriented programs (i.e., Direct Instruction and Behavior Analysis) to apply for funds. David Weikart (originator and sponsor of the Cognitively Oriented Curriculum Follow Through model) wrote to the Follow Through sponsors regarding the funding policy of ACYF:

You also notice that they (the federal government staff) ask for applications for the cognitive developmental approach and the developmental interactive approach. There is no RFP (Request for Proposals) requesting approaches on behavioral assistance. . . . They will not be requested in this round of RFP's (quoted in Carnine, 1983, p. 100).

ACYF claimed that there were no communities expressing an interest in implementing a behavioral approach. When a list of Head Start communities that did wish to establish a program based upon these approaches was forwarded to ACYF, they initially denied that the sites listed were official Head Start sites. Although it later recognized that any Head Start grantee could submit a proposal, no steps were taken to arrange for behaviorally-oriented models to apply for participation.

The goal of ACYF ostensibly is to develop effective educational strategies that help young children acquire developmentally appropriate educational skills. This goal appears to conflict with the actions of the agency. Not only did it fund programs known to be ineffective in teaching basic academic skills, it also favored the ineffective programs in its funding. These actions cannot be considered oversights. ACYF actively denied behaviorally oriented programs the opportunity to apply for funds. For example, ACYF never did issue an RFP for behavioral approaches. In other words, the federal government spent several hundred million dollars of tax money to determine the most effective methods for educating disadvantaged children so that rational decisions regarding federally funded compensatory education programs could be made. Yet, after the data were collected and the opportunity to provide funding for a compensatory education program came up, the methods most clearly identified as effective were excluded from consideration.

One final example of questionable use of federal research funds is offered. Mosteller (1975) suggested that a wise policy to follow in planned variation studies would be to identify the best looking programs and subject them to thorough testing. This would seem to be a reasonable course to pursue in Project Follow Through, particularly because of the fact that program components never were clearly specified. It is known for example, that the Direct Instruction and Behavior Analysis models had many common features. In fact, at one point the possibility of combining similar sponsors, such as these, had been considered. (This idea was rejected because it was recognized that despite their similarities, they also differed in some ways.) An interesting question is why the results of these two models differed despite their similarities. A component analysis would have allowed for far more than a demonstration that a "model" was effective. It would have

## USING THE RESULTS

provided information about the specific instructional components that were more or less effective.

This was not the course chosen by the Office of Education. During 1979, a team from the Office of OE's Assistant Secretary for Planning and Evaluation (ASPE) conducted an exploratory evaluation in order to develop new objectives, regulations, and directions for Project Follow Through (Wholey, 1979). It was agreed that 80% of Follow Through appropriations would be used to provide comprehensive services and 20% to fund research. In June, 1980, OE awarded four hundred thousand dollars of Follow Through funds (part of the 20% allotted to research) to the National Institute of Education (NIE) for the planning of a new series of Follow Through research and pilot activities. The agreement between OE and NIE called for NIE to provide a range of leadership and technical services to produce knowledge for use in improving the effectiveness of local Follow Through programs. The research direction appears to have been chosen primarily because of the finding of site variability. Rather than submitting the most effective Follow Through sponsors to a more thorough analysis, OE and NIE agreed to spend \$12 million to develop and study *new* Follow Through approaches. The intended purpose of this research was not, however, to identify instructional variables that were related to student achievement but to investigate management and implementation of educational innovations.

In 1980, OE awarded a contract to Abt Associates to conduct a search for potential new educational approaches and to assess the readiness of these approaches for implementation in Follow Through (Proper & St. Pierre, 1980). OE specifically requested that the search be limited to well developed approaches. Proper and St. Pierre stated that the primary concern was "whether or not an approach can be put in place and maintained, *not with the effectiveness of the approach in improving student outcomes*" [emphasis added] (p. 8).

The Office of Education seemed to be unaware that the two Follow Through models that were most easily implemented, Direct Instruction and Behavior Analysis, were also the most effective on measures of achievement (Stallings, 1975). If information about implementation was needed, these two models seemed to be a reasonable place to begin the analysis. The course pursued by OE shows almost total neglect of the findings of the Follow Through evaluation and a total disregard for academic achievement. It is conceivable that a model could be developed that could be perfectly implemented across a number of sites but that was ineffective in teaching academic skills to children. It seems obvious, but apparently was not to OE officials, that such a model would not be very useful or desirable.

The events described serve to illustrate the failure of results of the Follow Through evaluation to influence educational policies. While less formal documentation is available concerning the impact of Follow Through on educational practice, it is apparent that the Follow Through results have been largely ignored by the educational establishment (Greer et al., 1983). One cannot help but wonder why successful teaching methods have not been incorporated into American education. Why have the results of the Follow Through evaluation failed to impact the policies and practices of the educational community? Why have the most effective strategies for educating children



## *PROJECT FOLLOW THROUGH*

not been widely disseminated? Why has the knowledge gained from the Follow Through experiment not been utilized in the reform of education in this country?

The answer to these questions may be found in part in the manner in which instructional method is viewed by the various elements of the educational industry. Thus, in order to understand the failure of an industry to use these methods it is necessary to examine how the various elements of the industry view the interaction between students and teachers, the method of instruction. The remainder of this paper is an examination of sources of influence over the behavior of the various elements of the educational establishment with respect to the adoption of teaching methods.

*THE EDUCATIONAL ESTABLISHMENT*

## **PART V: THE EDUCATIONAL ESTABLISHMENT**

### **Policy Makers**

#### **Role of the Evaluation**

The purpose of evaluation is to aid in making rational policy and administrative decisions concerning public programs (Rossi, Freeman, & Wright, 1979). The rationale for the Follow Through planned variation experiment was to compare a variety of teaching methods in order to find out which methods were most effective in educating disadvantaged children. In other words, it was designed “to develop evidence to help guide policy decisions about the design and implementation of educational programs . . .” (Egbert, 1971, p. 1). Project Follow Through was intended to provide an empirical basis for decision making concerning federally-funded education programs.

When the Follow Through experiment began it was believed that policies would change in response to the identification of effective compensatory education approaches (Cohen, 1975). But in fact, policy makers and program administrators have not acted on the results of the Follow Through study. The data have not been used to make changes within the Follow Through program itself (Cohen & Garet, 1975) or in any other federally-funded education program. Policy analysts have observed that the politics of evaluation are such that decisions are frequently made without regard to the valuable and expensive information provided by the evaluation (Rossi et al., 1978). This seems to have been the case in Project Follow Through. The evaluation has been seemingly ignored in the decision making process.

Policy makers and administrators have not been sufficiently influenced by the findings to act on them either to direct further research or to alter existing compensatory education programs. Obviously, evaluation results are but one element in the decision making process. Other, more powerful sources of control may be found in the natural reinforcement contingencies existing within the government.

#### **Influence of Stakeholders**

Evaluation of public programs is always conducted in a context in which there are numerous parties with stakes in the continuation of the program. These stakeholders and their interests affect the way an evaluation is carried out. Examples of the way the interests of stakeholders influenced the conduct of the evaluation of Project Follow Through have been presented elsewhere in this paper. The decisions to add new sponsors and to shift the focus of the evaluation in response to demands of interest groups are just two examples of the influence of stakeholders over the conduct of the evaluation.

In addition to influencing the conduct of the evaluation, the interests of stakeholders affect the way results are used by policy makers and administrators. In fact, the influence of stakeholders appears to be considerably greater than that of evaluation results themselves (Rossi et al., 1979). One way stakeholders exert their influence is by generating controversy over the findings of the evaluation. Because field research typically lacks the degree of experimental control obtainable in laboratory investigations, social research is vulnerable to attacks on methodology. As the “perfect” natural experiment will never be done, Williams and Evans (1972) suggested that

## *THE EDUCATIONAL ESTABLISHMENT*

methodological flaws open the door to political issues and advised that in addition to the methodological substance of the criticisms the social concerns behind them be considered.

Indeed, much of the controversy surrounding Project Follow through has been cast in methodological terms. One of the most common complaints has been that outcome measures did not assess all of the goals of the program or of the individual sponsors. This belabors the issue. As discussed, the measurement of some goals is problematic both because measurement per se is difficult (for example, improved self-concept) and because the goals may be of questionable relevance to education. Failure to measure or address nonacademic goals does not reduce the importance or validity of the data obtained. Project Follow Through has been vulnerable to such criticism from the very beginning as administrators and policy makers failed to specify clearly the goals of the program.

When policy makers fail to specify goals, evaluators must determine what the intended goals of the policy makers were (or perhaps should have been). When objectives are not specified clearly, disagreement about evaluation findings is likely to occur. In the case of Project Follow Through, the goal was to “find out what works” in educating disadvantaged children. Because these children are characterized by their poor performance on traditional measures of educational achievement, and because these tests present a fair sample of academic skills, these tests seemed a reasonable basis for comparing the effectiveness of a variety of teaching approaches. In fact, the selection of the instruments used in the test battery represented a consensus among parents, educators, taxpayers, and legislators about important objectives of schooling (Rhine, 1983).

Yet, following the evaluation, the choice of outcome measures became the focus of a great deal of controversy. Stakeholders have countered the negative findings of most approaches with the claim that these models were effective in areas not included in the outcome measures of the evaluation. In addition, individuals with a stake in models that did not fare well on the evaluation have asserted that goals that were not measured are of greater importance than those that were included in the evaluation (that is, academic achievement).

It seems inevitable that some party or special interest will be dissatisfied with any evaluation. It seems obvious, however, that ultimately educational approaches should be evaluated primarily on the extent to which educational achievement has been affected. This was the basis for original inclusion in the experiment and the area of critical importance for disadvantaged children. This goal of improved academic skills is not inconsistent with the more global goal of “improving life chances” that is frequently cited as the goal of compensatory education. The skills measured in the Follow Through evaluation are essential for effective functioning in this society. Improved life chances may be a consequence of attainment of academic skills, but is too vague to serve as an objective of evaluation. Nevertheless, the claim that certain approaches were effective in areas that were not, or could not be, measured cannot be disputed. As no data are available, no statements or comparisons in these areas can be made. Thus, the failure of policy makers and administrators to specify objectives made it difficult to obtain a commitment to accept the evaluation findings (Williams, 1972).

## *PROJECT FOLLOW THROUGH*

The concern over methodology waxes and wanes in response to political concerns. Recall that the Wolff and Stein (1966) study was credited with establishing the very need for Project Follow Through despite its serious methodological inadequacies. Yet, ten years later, when the results of the evaluation of Project Follow Through became available, methodological shortcomings were offered as justification for the failure of policy makers and administrators to act on the results. In field research, all evaluation studies have some methodological frailties, and those who take issue with evaluation results have been diligent in finding methodological experts to attack an evaluation (Rossi & Williams, 1972). The scathing critique by House et al. (1978) is an example of this practice with respect to the evaluation of Project Follow Through. Yet, despite acknowledged methodological weaknesses, the methodological sophistication of Project Follow Through far surpasses that of any other federal educational evaluation. When weak methodology is opportunistically invoked as a rationale for the neglect of research findings, it can be concluded that methodological rigor becomes an issue only when the evaluation results do not support the interests of a majority of stakeholders.

The imperfections in the experimental design of Project Follow Through were recognized long before the evaluation results became available. They were there from the inception of the experiment. Yet, there had been little concern expressed over the effect these shortcomings would have on the results. In fact, the expectation was that the effect of the program would be so great as to overcome any weakness in methodology. Former OE Project Director, Mary Kennedy (quoted in Benjamin, 1981) charged that although the flaws were known before the study began,

... it wasn't until the findings came in that the methods were criticized. Even when you correct the flaws in the study though, the results turn out exactly the same. There were just a lot of academics and liberals who favored the idea of "open" education and didn't like it when the structured programs won out and the "open" models looked bad; that's when the criticisms started. (p. 91)

The fact that concerns about methodology were generally not expressed until after the results of the evaluation were available suggests the likelihood that interested parties would have accepted the methodology as adequate had the findings favored their interests. This surmise is certainly supported by the failure of these critics to attack the results of the original Wolff and Stein study.

Policy makers did make decisions about the fate of Project Follow Through. The program was not automatically renewed year after year, but funds were actively appropriated each year. The transcripts of various hearings before Senate and House committees include testimony from numerous stakeholders. The general message conveyed to the members of Congress during these meetings was that all of the educational methods included in Project Follow Through were successful. The fact that the data do not substantiate this claim has been countered with charges of inadequate methodology. If, however, the experiment was so poorly designed as to make the primary findings unacceptable, no other conclusions can be drawn from the study either. Any claims that other models were effective, or claims that effects have taken place in outcomes other than those included in the evaluation are pure conjecture. Claims of effectiveness must be substantiated with data. The fact that the mere assertion of

stakeholders that all Follow Through models were effective appeared to be sufficient to control the behavior of policy makers attests to the fact that public policy is based on public support, not on empirical evidence.

The influence of stakeholders is clearly substantial, but it is not the only factor to determine policy decisions. Another important source of control is the expectations of the policy makers themselves.

### **Expectations**

Rossi & Williams (1972) cited administrative expectations as an important source of control over the use of evaluation findings. In fact, some social policy analysts assert that in situations in which administrators are strongly convinced of the effectiveness of a program, it is likely that an evaluation will be disregarded. This observation has important implications for the use of the evaluation results of Project Follow Through.

It is apparent that although protests over the Follow Through evaluation are often framed methodologically, interpreting the results in policy terms involves considerations that go well beyond methodology. In addition to the influence of their constituents, decision makers are influenced by their own expectations. The term expectations, as it is used here, is defined as “predictions about the outcome of an evaluation” and these predictions are a function of an individual’s conception of the process of education. Expectations played an important part in Project Follow Through. The expectations of planners and administrators exerted considerable control over the design of the experiment and evaluation, as well as the reporting of the results and the eventual use of the results by policymakers.

The methodological weaknesses of the Follow Through experiment were the result of a number of factors, among them the expectation of program administrators that the impact of the Follow Through approaches would be so great that the effects would overwhelm any weakness in design. This assumption was expressed by Egbert (1973) in the following statement:

[The] design [of Project Follow Through] stemmed from the conviction that sufficient improvement could be affected in the institutions serving children that children’s development would be so markedly superior as to be readily demonstrated on measures of achievement, cognition, self-concept, social maturation, and capacity to function independently. . . . In view of the results reported by Miller, Engelmann, Gordon and others in the January, 1968, meetings of prospective sponsors, this conviction did not seem unrealistic . . . (p. 25)

The above statement reflects the naiveté of the administrators about the complexity of the task of improving academic performance of disadvantaged children. It had been well established that disadvantaged children did not benefit from traditional educational approaches. The results of evaluations of federal compensatory programs suggested that compensatory education programs did not result in any significant advantage over traditional classroom teaching. Nor did there appear to be any evidence in the sponsor presentations that would have caused one to expect a “slam bang” effect. The results presented by the sponsors at the meetings in January, 1968, demonstrated that in some cases gains on certain measures could be obtained in highly controlled laboratory

## PROJECT FOLLOW THROUGH

schools. There was little reason to believe that the same degree of control could be obtained in the field and no reason to expect that even more substantial gains would result. Certainly, there was no basis for expecting results so robust that methodology could be ignored.

Egbert's statement also reflects the ignorance of administrators and planners about the importance of instructional method as the primary determinant of learning. Note that the conviction was that institutional improvement would lead to superior student outcomes. It is not clear what institutional changes Egbert considered to be related to student achievement. It is clear, however, that he did not consider achievement to be a function of the method of instruction, which was the independent variable of the experiment. Instead of looking for relationships between teaching method and student outcomes, that is, differences between models, the Follow Through staff hoped to see differences between students participating in Follow Through and those who did not. Thus, while the planned variation approach was intended to reveal differences in effectiveness among different teaching methods, early evaluations focused on features common to all models in order to justify the question of overall Follow Through versus non-Follow Through comparisons (Haney, 1977b).

The expectation that program effects would be great influenced the manner in which results were reported. Alice Rivlin (1971) discussed government interference in reporting of results and asked ". . . where the experimental results are mixed, might there not be a tendency to emphasize the positive results while deemphasizing, if not actually suppressing, the negative ones?" (p. 113). But what are "positive" and "negative" results? The outcome of an experiment can be judged positive or negative only in terms of its agreement or disagreement with a prediction or expectation (Sidman, 1960). Given the expectations expressed by Egbert, results would be judged positive only if they indicated overall program effectiveness.

To see how this expectation influenced the reporting of the results, recall the decision made following the release of SRI's first year report. When the results of the first year evaluation became available, it was clear that the data did not support the hypothesis of Follow Through planners and administrators. Consequently, "program administrators and evaluators [*sic*] deliberately buried the evaluation results. . ." (Elmore, 1976, p. 374).

A second example will serve to illustrate not only the tactic of emphasizing "positive" findings in accordance with expectations, but also the influence this practice has on the decision making process. Although Abt's initial evaluation efforts continued to pursue the question of overall program effectiveness, in their first report (Cline et al., 1974) it was concluded that model effects were not homogenous and that "structured" programs performed significantly better on some measures. This finding of differential effectiveness was nowhere apparent in the summary of the draft of this report that was prepared by OE. At an appropriations hearing, Senator Norris Cotton read from the OE document "when contrasting all Follow Through children with their non-Follow Through comparisons across all sponsors in all parts of the country, there emerge large differences in achievement, motivation, and intense effects" (U.S. Congress, 1974, p. 2361). Senator Cotton was convinced that the Follow Through program had been successful and should be continued.

## THE EDUCATIONAL ESTABLISHMENT

John Evans, then OE's Acting Deputy Commissioner for Planning, Budgeting, and Evaluation, however, questioned whether it makes sense at all to conclude that Follow Through was purely a success or failure in one single phrase that way. Evans went on to explain to members of Congress:

. . . Follow Through is made up of a different set of alternative ways of approaching compensatory education, different models, different programs. And the task and central purpose of that program, which is, . . . primarily an experiment, is to find out which of those methods or approaches are more or less effective. The evaluation evidence we have compiled indicates just what we would expect from that kind of experiment: namely, that some of those models and approaches are very reassuringly effective, and the kinds of things we would want to see disseminated and used more broadly in compensatory education. But that same evaluation evidence indicates that a number of the other models are not successful and not effective and not the kinds of things that we would want to carry on or continue to fund or support indefinitely. (U.S. Congress, 1974, p. 2360)

Reports of overall program effectiveness served as a source of confusion and controversy when decision makers were faced with the task of determining the fate of the program. The above example indicates the manner in which expectations influenced the conduct of the evaluation question and the reporting of findings. In turn, as policymakers rely on information provided them, their decisions are based often on incomplete and inaccurate data that reflect not what the experiment has revealed, but the biases of program administrators.

In summary, there are two essential questions to ask about policy makers' disregard for the findings of Project Follow Through. The first question is whether the results were sufficient for directing policy decisions. Some (e.g., Elmore, 1976) have argued that the Follow Through experiment does not yield the kind of information necessary for policy decisions. However, the results of the Follow Through experiment, at least with respect to measures of student achievement, are quite conclusive. All educational approaches were not equally effective. The experiment indicated which methods produced the greatest gains for disadvantaged children.

The second question that must be asked about policy makers concerns their original intent about how they would act on the findings. In reference to the early planning meetings, Alice Rivlin (1971) stated:

While hardly anyone thought it was appropriate for the federal government to dictate what curricula or methods ought to be used in local schools, even under federally funded programs, there was strong support for federal efforts to help communities to make more informed choices by fostering planned variation experiments and *making results widely available* [emphasis added]. (p. 94)

Mary Kennedy, former OE project director, provides a different perspective. Kennedy (quoted in Benjamin, 1981) claimed that due to the extreme sensitivity to anything that smacks of a national curriculum coming down from the federal government, "no one in Follow Through from the start really expected the government would do anything about the study" (p. 32).

Elmore (1976) presumed that OE must have been uncomfortable having invested millions of dollars only to find the data inconclusive. Careful consideration of the data



## *PROJECT FOLLOW THROUGH*

and the contingencies affecting policy makers and administrators suggest an alternative interpretation that any discomfort experienced by OE was occasioned not by a lack of conclusive data but by the nature of the findings. The two models that demonstrated significantly superior performance on academic measures, Direct Instruction and, to a lesser degree, Behavior Analysis are unquestionably outside of the mainstream of American education. Considering this, it is not surprising that the evaluation findings have raised considerable consternation in the educational community.

Protests over the interpretation of the data were to be expected in an experiment in which so many people had so much to lose (and to gain) based upon the findings. In a program of the magnitude of Follow Through the stakes were enormous. The consequence of these criticisms and cries of “unfair” has been a retreat by federal educational agencies to a neutral position. The influence of stakeholders has resulted in restraint in the reporting of results and tempering of the impact of the findings.

While it is acknowledged that policy makers are more likely to be influenced by social and political contingencies than by empirical data, others involved in education may be expected to pay more heed to the findings of major research programs in their profession. Some (Lyons, 1980; Skinner, 1984) have suggested that the blame for the failure of American education should be placed on colleges of education. Reasons for the failure of schools of education to use the findings of Project Follow Through will be discussed next.

### **Colleges of Education**

#### **Conflicts With Existing Philosophies**

Project Follow Through was unique because it compared the effectiveness of instructional methods that not only differed from one another but that were each consistent with a particular set of beliefs about learning and education. Project Follow Through examined not only teaching methods but the educational philosophies from which the methods were derived. In addition, it was the first major attempt to do so.

While the Follow Through models varied greatly in specific differences, they may generally be considered to represent one of two general philosophies of education. The majority of models were based on philosophies of “natural growth” (Becker & Carnine, 1981), or what Bijou (1977a) has referred to as “unfolding.” According to these models, learning involves changes in cognitive structures that are believed to develop and mature in the same manner as biological organs. The second philosophical position was concerned with principles for “changing behavior” in desired ways (Becker & Carnine, 1981). From this perspective, teaching involves specifying what is to be taught and arranging for that change in behavior to occur.

The results of the Follow Through evaluation indicate that the models based on the latter philosophical position were the most successful in improving academic achievement. The meager representation of this point of view indicates the relative popularity of this approach in education. The majority of colleges of education espouse educational philosophies that are based on a philosophy of cognitive restructuring (Skinner, 1984). The results of the evaluation of Follow Through strongly suggest that such philosophies are not conducive to the development of effective teaching methods.

## THE EDUCATIONAL ESTABLISHMENT

The Follow Through models that approached teaching from this perspective were generally unable to improve student performance significantly on measures of basic academic skills. The two models that did result in a significant impact on academic performance, Direct Instruction and Behavior Analysis, were both founded on the premise that academic skills must be taught directly. Thus, the Follow Through data fail to support the philosophies of education that have been adopted by the majority of professional educators. Educators have responded to the findings by presenting arguments in defense of cognitive philosophies. These efforts have been aimed at discrediting the findings either by criticizing the national evaluation in general (e.g., House et al., 1978) or by voicing specific objections to the Direct Instruction and Behavior Analysis models. Examples of these strategies follow.

Proponents of philosophies of cognitive restructuring concern themselves with issues that they presume relate to long lasting effects of education (Becker & Carnine, 1981). Some educators have contended that the Follow Through evaluation focused exclusively on immediate and short term effects and made it appear that the unstructured methods were ineffective. Failure of these models to obtain results on achievement tests is justified by the claim that the effects are delayed. From this point of view, failure of the majority of Follow Through models to improve achievement scores significantly should not be taken as evidence of the ineffectiveness of those models. According to this view, instruction has resulted in changes in underlying abilities or cognitive structures that are not readily apparent or easily measured. Edwards and Bridewell (1979) articulated this position in the following statement, “a primary student who is deficient in basic skills is no tragedy if his expectation for the later acquisition of such skills has been improved as a result of educational experience” (p. 8). This opinion is also expressed in the Responsive Education model’s belief that

... in order to learn, a child needs a sense of personal worth and an environment in which materials and activities stimulate and respond to the child’s interests. Given that essential self-esteem and that learning environment, *the acquisition of academic skills will follow*” [emphasis added]. (Stebbins et al., 1977, p. A-79)

In order to substantiate the claim that instruction will result in benefit at a later date, it is necessary to provide evidence of achievement at some point following the conclusion of the intervention period. Some researchers (e.g. Riley, 1978; Seitz, Apfel, & Efron, 1977) have evaluated the long term effects of unstructured models. These studies were designed to determine if Follow Through models that aim for achievement gains by first influencing mediating variables produce delayed or “sleeper” effects. After reviewing several studies of delayed effects, Goodrich and St. Pierre (1979, 1980) concluded that there was no evidence of any sleeper effects. That is, academic improvement was not seen in the students who participated in those Follow Through models, even when the evaluation of academic skills was conducted 3 or 4 years following intervention.

Educators may discredit the Behavior Analysis and Direct Instruction models by questioning the values these models represent. Educators are fond of accusing direct teaching approaches of ignoring “the whole child” by emphasizing achievement at the expense of affective development (Lyons, 1980). Affective development, the focus of

## *PROJECT FOLLOW THROUGH*

many Follow Through models, is undoubtedly of great importance. However, educators who have focused on affective development as the primary goal of education have generally not been able to demonstrate that the activities in which they engage do in fact result in improved self image (Novak, 1977). Even if they could, it would be questionable whether the gains in self-concept would compensate for the resulting academic deficits. The Follow Through evaluation provides clear evidence that no such trade off is necessary. The Direct Instruction and Behavior Analysis models not only resulted in superior performance on achievement measures, they were ranked first and second, respectively, on measures of affective development. It appears that the way to improve the self concept of children is to provide them with successful experiences by programming instruction in such a way that the child can see evidence that he or she is learning.

These are just two examples of the way in which professional educators have responded to the challenge of their philosophies that the Follow Through results constitute. It should be noted that the effectiveness of the Direct Instruction and Behavior Analysis models generally has not been questioned. The reaction of schools of education to the results of the Follow Through evaluation suggests that educators judge instructional methods not by their effectiveness, but by their congruence with prevailing philosophies of education. This should come as no surprise, however. According to Skinner's (1957) analysis, we would expect that the tendency to act on the Follow Through findings would be identical with the "beliefs" of professional educators. In other words, because these individuals do not share the same philosophical views as the developers of the Direct Instruction and Behavior Analysis models, the results of the evaluation have failed to alter their behavior. The possibility must be considered, however, that professional educators have failed to act on the findings of the evaluation of Follow Through for more practical reasons as well.

### **Inadequate Skills**

Providing teachers with training in those methods that have been identified as effective would entail restructuring of teacher training programs. These changes would involve not only what teachers are taught but how they are taught. Such changes are unlikely to occur because those who train teachers lack the necessary repertoire as well as the resources for acquiring such skills.

Professionals who are trained in Behavior Analysis and Direct Instruction have some difficulty obtaining appointments in traditional departments of education (Greer, 1982). Consequently, the individuals who have the skills to train teachers effectively are least likely to contact those prospective teachers. Even if this situation were to change, the number of professionals who are proficient in the use of direct teaching methods is so small relative to the number of students enrolled in colleges of education that they could not provide the needed training for all teachers.

In order for all teachers to become competent in the use of effective teaching methods, those who are currently responsible for teacher training must themselves become proficient in the use of those teaching strategies. But, how will they acquire such skills in the absence of sufficient numbers of competent trainers? More importantly, even if those responsible for teacher training could become sufficiently familiar with the

techniques of the Behavior Analysis and Direct Instruction models, they would be unable to transmit the methods to teachers using the training paradigm found in most colleges of education.

The training paradigm underlying most teacher training programs has little to recommend it. Teacher trainees spend the majority of their time listening to lectures about theories of learning or about what teaching should be. Sponsors of both structured and unstructured Follow Through models found that instruction in the form of lectures about theory and method had little impact on teaching practices (Weikart & Banet, 1975; Bushell, 1978). The sponsors of the Behavior Analysis Follow Through model found that teacher behavior was altered and maintained only when training provided modeling of the desired behavior, opportunities for the teacher to engage in the behavior, and feedback about his or her performance (Bushell, 1978). The conditions under which teachers acquired teaching behavior were essentially the same as the conditions specified by the behaviorally-oriented models as essential for learning to occur with children. In other words, the most effective way to train teachers to use effective teaching methods is to apply the principles on which those methods are based to the teacher training paradigm.

This finding has profound implications for those who are responsible for teacher training. It suggests that in order to train prospective teachers in methods that are effective in changing the academic behavior of their students, professors of education must acquire effective teaching skills and use them to train teachers. To do so would likely require that professors of education seek out this training. While an occasional professor or two may go to the time, trouble, and expense required to become skilled teacher trainers, the contingencies of reinforcement to which these professionals are subject generally do not support such activities.

### **Lack of Effective Contingencies**

We have learned from the experimental analysis of behavior that if desirable behavior is to be strengthened, that behavior must be reinforced. Thus, if effective teacher training is desired, those professional activities of college instructors that are most likely to result in effective teachers should be reinforced. Unfortunately, professional rewards and recognition are generally not arranged for effective teacher training. The lack of value placed on teacher training in colleges of education may be explained by the fact that teaching is generally not regarded as a set of skills acquired through training, but as an art in dealing with people. In other words, professors of education are not rewarded for turning out competent teachers because good teaching is assumed to be a function of factors other than professional training.

Similarly, schools of education may not be held accountable for teachers who are not effective because this deficit is also attributed to some deficit in the teacher, not in the training provided by the teacher training institution. The notion that teachers are born and not made absolves professors of education of the responsibility of making sure that teachers enter the classroom with the skills necessary to teach academic skills to children. An additional effect is that professors of education must look elsewhere to obtain professional reinforcers.

Reinforcement in the form of salary increases, tenure, and promotion is more likely to be arranged based on number of publications or attainment of outside funding

## PROJECT FOLLOW THROUGH

than on training teachers. Such a contingency is not likely to result in professional educators using valuable time either acquiring effective teaching skills or engaging in the behavior needed to train teachers effectively. It is more likely that the job of teacher training will be given cursory attention, with the majority of the professors' time being devoted to research.

This practice would not be so counterproductive to the goal of improving education if research efforts were primarily devoted to analyzing and delineating effective teaching methods. But again, the contingencies do not favor these activities. Publications and grant money are more likely to be awarded to those who pursue educational questions from a cognitive orientation. For example, Greer (1982) provided evidence that research articles are more likely to be published by leading journals in the field of education if the research is aimed at cognitive restructuring. Behaviorally-based research is rarely published in leading educational journals such as the *American Educational Research Journal* and *Educational Researcher*.

Funding agencies have also favored cognitively-oriented research. The National Institute of Education's (NIE) 1980 guidelines for solicited research grants expressed a strong preference for studies of cognitive processes (Greer, 1982). In the late 1970's NIE turned down five research proposals from the developers of the Direct Instruction model (Benjamin, 1981) after it was determined to be the most effective Follow Through model. Thus, research that is behaviorally-based is less likely to be published and/or funded than research on cognitive processes. Yet, professional reinforcers are largely contingent on obtaining publications and grant money. The net effect is an abundance of research that the American Educational Research Association (cited in Greer, 1982) has acknowledged has not yielded useful instructional procedures. Despite the apparent failure of both the research and teacher training activities of colleges of education, these institutions have remained largely unaccountable to anyone for their practices.

### **Publishers**

#### **Economic Contingencies**

The actual materials used in instruction may be distinguished from the method of instruction. Bushell (1978) argued that instructional method, or pedagogy, is the critical factor in ameliorating learning problems. The majority of Follow Through sponsors used curriculum materials that were available commercially. The one notable exception was the Direct Instruction model of the University of Oregon. The sponsors of the Direct Instruction model developed their own instructional materials based on complex and sophisticated programming principles (Engelmann & Carnine, 1982). The finding that the Direct Instruction model's performance was superior to all other models suggests that curriculum may be more intimately related to student outcomes than previously believed and has important implications for the design of educational materials. To the extent that instructional materials control the behavior of the teacher, they are partly responsible for student achievement. The failure of publishers to incorporate knowledge about effective instructional practices into their products contributes to the continuation of ineffective teaching practices.

## *THE EDUCATIONAL ESTABLISHMENT*

It is not surprising that the results of the Follow Through experiment have not affected the practices of the publishing industry. To see why this is the case one need only examine the contingencies governing the publishing industry. These contingencies are economic rather than academic. Unlike the educational establishment in general, the publishing industry has a clear goal, to make a profit. In order to make a profit, they must respond to the preferences of the consumer. For this reason, publishers do not generally react to the findings of educational research but to the interests of the educational community.

The interests of the educational community are usually represented by the publishers' advisory boards, consultants, and authors, all of whom are professors of education. The negative reaction of this group to direct teaching methods has already been discussed. The influence of professors of education is pervasive. The approach to education that is transmitted through the colleges of education does not establish the need for carefully designed instructional materials. Thus, professors of education are instrumental in creating a situation in which it is unlikely that publishers will produce carefully programmed instructional materials.

Rather than promote teaching as a technology, college professors have subscribed to the notion that curriculum materials must be individualized and that curriculum must be matched to the needs of the learner. While the Follow Through results indicate that individualization does not require a plethora of materials or methods, publishers are unlikely to respond to this finding. The publishing industry has benefited greatly from the concept of individual differences. Any broad consensus by educators about teaching practices and materials would no doubt result in a financial drain on the publishing industry.

While educators encourage diversity, these variations must fall within a narrow range. Anything that is too great a departure from traditional teaching practices will not be likely to be adopted. As Carnine (1984b) explained, "Publishers do not readily incorporate knowledge about effective instructional practices into their textbooks for fear of creating an unusual and unsuccessful product that might reduce their share of a multi-million dollar market" (p. 19). There is no doubt that the instructional materials developed by the sponsors of the Direct Instruction Follow Through model are unusual in the sense that they do not resemble traditional textbooks. Direct Instruction materials are published by Science Research Associates. Those materials make up only a small percentage of the total sales of that company. The primary reason cited for the dearth of sales of Direct Instruction materials is that the materials do not resemble closely other texts and workbooks (W. C. Caruth, personal communication, February 18, 1986).

Obviously, educators do not evaluate instructional materials based on their demonstrated effectiveness. Lovitt (1977) described the way in which curriculum materials are evaluated as "topographic and subjective." Publisher's descriptions of their materials are replete with topographical descriptions of their products. Sales catalogs reveal that the characteristics of the materials most frequently described have to do with the dimensions of the book or kit, the number of pages, the number of color prints, etc. Conspicuously absent from these sales pitches are data related to effectiveness of the curriculum. Because publishers continue to sell instructional materials in the absence of

## *PROJECT FOLLOW THROUGH*

data to support the claims of the publisher, it must be assumed that empirical evidence of a program's effectiveness does not greatly control purchasing by school districts.

The National Commission on Excellence in Education (1983) recently requested that publishers provide evidence of the effectiveness of their products based on field research and evaluation. However, there is little incentive for publishers to respond to this request. Conducting field trials and evaluations is an expensive endeavor, one from which the publishing industry is unlikely to derive much benefit. There is little reason to believe that the educational community will attend to data provided by publishers any more than they have to educational research on teaching methods such as Project Follow Through.

In general, the publishing industry is passive. They do not conduct research or respond to available research about instructional practices. They do not initiate the development of new products because it is financially risky to do so. Instead they respond to the consumers of educational materials. The effectiveness of a product is measured not by the extent to which it results in academic achievement, but by sales. The educational community has failed to adopt the approaches identified as effective in the Follow Through evaluation and thus has created no substantial market for Direct Instruction materials.

### **Instructional Designers**

Even if the demand were generated, it is difficult to see how the publishing industry might respond to it. The developers of the Direct Instruction model found it necessary to develop their own instructional materials because they determined that there were no commercially available materials that were adequately constructed to yield educational gains from disadvantaged children. The programming principles embodied in the Direct Instruction materials are extremely complex and much more sophisticated than that of any other available curricula. The number of individuals who are proficient in designing these types of materials is quite small. Furthermore, these individuals are not employed by publishing companies.

In order for publishers to produce materials based on Direct Instruction programming principles, they would need to find and hire individuals with those skills. That is likely to be a difficult task. An alternative would be to contract with individuals who are skilled in the development of such materials. This alternative would not be very attractive to publishers because contracting for those services would reduce the publisher's margin of profit.

There is, of course, another alternative: to publish materials that resemble Direct Instruction materials but are developed by authors with inadequate understanding of the instructional design principles upon which Direct Instruction is based. This is what happened with programmed instruction materials (Skinner, 1963). The demand for instructional materials is a double-edged sword. If educators do not demand Direct Instruction materials, they are unlikely to be developed and produced. If there is a great demand, publishers are likely to produce materials that are of poor quality because those who author programs are unlikely to be adequately trained in instructional design.

The persons who develop instructional materials are often teachers or those who have previously been teachers. It should not be surprising, then, that educational materials reflect more traditional notions of instruction than those that were used by the

most successful Follow Through sponsors. Many of the textbooks available for purchase are written by individuals with no formal training in instruction and no experience in teaching (Carnine, 1984b). It is difficult to see how these individuals could develop the carefully programmed instructional materials that lead to effective teaching.

## **School Districts**

### **Financial Contingencies**

There was a consensus among government officials that it would not be “appropriate for the federal government to dictate what curricula or methods ought to be used in local schools, even under federally financed programs” (Rivlin & Timpane, 1975, p. 2). It was agreed, however, that the government should assist local school districts in making informed decisions about curricula and methods by conducting planned variation experiments, such as Follow Through, and by making the results widely available. The government’s policy of making results available assumed that local school administrators were motivated to look for better technologies and that if information about effective practices were available these methods would be adopted. The assumption that districts would select the most effective methods has not been supported by actual practice, however. School representatives generally do not seek out effective teaching methods or adopt them when they are available (Berman & McLaughlin, 1975).

In short, the mere fact that effective teaching methods are available does not guarantee that they will be adopted by school districts. Even when districts adopt Follow Through models, it has frequently been those methods that failed to be effective in improving student performance on achievement measures. This appears to be either because there are no incentives for adopting effective methods or because of penalties for adopting ineffective methods (Pincus, 1974). In the absence of explicit contingencies to select methods based on demonstrated effectiveness, the choice of teaching methods will be influenced by other factors.

One reason administrators may adopt a particular method is because it is financially profitable to do so. Benjamin (1981) reported that certain urban school districts operate up to 60 different compensatory education programs, each with its own source of funding and method of instruction. The primary goal does not appear to be attainment of effective services for students, as much as the acquisition of money as an end in itself. This goal is perhaps related to the misguided belief that money will result in improved education. It is not difficult to identify conditions that gave rise to this line of reasoning. Because education is largely funded by local taxes, poor communities have less money to allot to schools. Because poor children who attend poor schools typically perform at a lower level than their more affluent peers, it may be assumed that increasing the amount of money available at a given school will result in educational opportunities more commensurate with that of more affluent schools.

The results of the Follow Through evaluation dispel the notion that additional money is sufficient to improve educational performance. All Follow Through projects were funded at the rate of about \$350 per child per year above the basic school support for the educational component of the program (Bushell, 1978). Despite the fact that all sponsors had about the same amount of money with which to provide an educational



## *PROJECT FOLLOW THROUGH*

program, the results indicate that improved academic performance was a result of the type of instruction provided. In other words, money alone was not sufficient to increase achievement (Becker, 1978), and it cost no more to implement effective than ineffective approaches.

In some cases, financial contingencies may actually make it undesirable for administrators to adopt effective methods. For example, if distribution of money is based on low achievement, then any intervention that increased student achievement would result in a decrease in funding. Michael O'Keefe, Deputy Assistant Secretary for Planning and Evaluation in HEW, expressed this contingency when he cautioned that "allocating money on academic criteria might serve as a disincentive to local districts in that the higher the achievement of their students the less money they would receive" (OECD, 1980, p. 104). Benjamin (1981) alleged that some low income schools using Direct Instruction actually lost special funding when student achievement increased significantly.

In general, teaching methods are poorly controlled even in federal programs. Financial contingencies are seldom enforced. For example, in Project Follow Through school districts agreed to adopt the teaching methods of a sponsor as a condition of participating in and receiving funds from Project Follow Through. However, despite the fact that in one site the school district continually failed to cooperate in implementing the model, leaving the sponsors no alternative but to withdraw their services, the school continued to receive Follow Through funds. The situation may be summarized in the following way. When money is offered as a condition for adopting an innovation, school districts are likely to do so, at least nominally. In the absence of financial compensation there is little incentive for school districts to adopt an innovation. In fact, when asked before the House of Representatives appropriations committee whether there was any indication that local school districts would use non-Federal monies to support Follow Through models, Assistant Secretary of Elementary and Secondary Education Larry Davenport responded that there was no evidence to suggest that they would.

### **Local Norms**

If effectiveness is not the primary determinant for choosing educational innovations, what is? Instructional decisions are most likely to be controlled by variables unrelated to student outcome. The majority of professional reinforcers available for principals and central administrators are contingent not on student performance, but on the continued and relatively smooth operation of the school.

Consequently, those innovations that are likely to result in the greatest amount of disruption are the least likely to be adopted. Thus, whether a particular method is adopted by a school district depends on the extent to which the innovation differs from current practices (Gaynor, Barrows, & Klenke, 1980). The most likely candidates for adoption are those approaches that are most similar to ongoing practices. However, because existing practices are generally not effective for teaching academic skills, it may be assumed that the more an innovation resembles existing practices, the more similar it will be in its effects. In other words, the methods that are most likely to be adopted are the least likely to improve student outcomes.

## *THE EDUCATIONAL ESTABLISHMENT*

The probability that the instructional methods employed in the Direct Instruction or Behavior Analysis Follow Through models will be readily adopted by local school districts is fairly remote. Lloyd Cooke said in reference to the Direct Instruction model, "The problem with the program is that it violates everything that's been held up as tried and true in education. It requires so much change in educators' behavior that it almost can't succeed in most school systems today" (quoted in Benjamin, 1981, p. 90).

### **Organization**

Project Follow Through provided a unique opportunity for educators to examine the conditions necessary for implementing teaching approaches in school districts. In the case of most educational reforms, the developer of the innovation is not responsible for its implementation. In contrast, Follow Through sponsors had the responsibility of making sure their curricula and methods were enacted (Beers, 1976). Although the sponsors differed in the manner in which they approached this task, they have, for the most part, been successful. While precise implementation measures are not available, the observation studies that were contracted (e.g., Stallings & Kaskowitz, 1974) indicated that there were differences in actual classroom practices reflecting the various models. Weikart and Banet (1975) expressed the opinion that school systems would not have been capable of fulfilling the function of the model sponsors. In other words, school systems, as they are currently designed, cannot support the implementation and long term maintenance of educational innovations. In general, school districts lack adequate teacher training and systems for monitoring and providing feedback.

Tomlinson (1981) described academically successful schools as organized around the goal of student learning. This is consistent with Clark, Lotto, and McCarthy's finding (cited in Carnine, 1984b) that successful low income schools are characterized by clearly stated goals and objectives. These researchers also concluded that the more specific the goals of a program, the more likely the program is to be successful. These observations were supported by data from Project Follow Through. Based on observations of Follow Through classrooms, Stallings and Kaskowitz (1974) concluded that those programs that were most successful were those that had clearly specified objectives.

The general implication of these observations is that effective schools are product oriented. They are characterized by agreement from the top level administrators down to the classroom teacher that the desired outcome is student achievement. All decisions are then based on the degree to which they are effective in reaching that goal. For example, the administration might elect to adopt teaching methods that have been demonstrated to be effective, provide adequate staff and teacher training, and monitor teacher performance to ensure accurate implementation and maintenance of the method. This, of course, is based on the assumption that teaching method is the primary determinant of student achievement.

School districts, however, are generally not organized in this manner. While at first glance school districts seem to be hierarchical in organization, with educational practices determined by administrators, implemented by teachers, and supervised by principals, in reality the members of a local school system (teachers, principal, and central administrators) behave in ways quite independent of one another (Morris,

## *PROJECT FOLLOW THROUGH*

Crowson, Hurwitz, & Porter-Gehrie, 1981). In other words, although administrators are commonly considered to be instructional leaders, they do not function in that capacity. In general, administrators are more likely to be concerned with the successful maintenance of the school as an organization, focusing on teacher credentials, student selection, physical plant, and finances (Meyer, Scott, & Deal, 1979). Decisions regarding the method of instruction are poorly controlled, if managed at all, by administrators of local school districts.

### **Monitoring and Feedback**

Even in cases in which administrators decide to adopt a teaching method that has been demonstrated effective, initial training in the method alone is seldom sufficient. One important lesson learned by Follow Through sponsors was that workshops were insufficient for bringing about the type of behavior change required (Bushell, 1978). Yet, a one-day, or perhaps weekend, workshop is the typical form of inservice training provided by schools. This practice may be attributed to a cognitive conception of learning. It is based on the assumption that knowing about methods is equivalent to knowing how to teach. One thing that became obvious to all sponsors was that training needed to focus on teacher behavior. Implementation refers to behavior. It can be said that a teacher “understands” the method only when he or she demonstrates its appropriate use. In the absence of this demonstrated behavior, we have no way to evaluate whether the teacher “knows” how to use that teaching method. The experiences of the Follow Through models suggest that the continued use of workshops and inservice training by school districts will not result in significant changes in teacher behavior.

Another critical component for successful implementation is a mechanism for monitoring and correcting teacher performance. Monitoring and correction are common in industries that are product oriented, but are relatively unused in education. One effect of this lack of monitoring is the continuation of inefficient teaching practices. It is not unusual for even the most effective teaching method to become distorted in practice without careful and systematic monitoring. Keller (1985) explained that new teaching methods are frequently distorted to the point that teaching is essentially unchanged from that occurring prior to introduction of the innovation.

Part of the explanation for the failure of school districts to establish controls such as these over instructional practices is the way they view teaching. As David Cohen (quoted in Carnine, 1984b) explained “If teaching is essentially an intuitive and unknowable activity, then there is no way to control teaching acts through monitoring of either the methods used or the outcomes achieved” (p. 16).

The Follow Through models that were most effective in raising achievement scores, Direct Instruction and Behavior Analysis, were also the most easily implemented models (Stallings & Hentzell, 1978). This finding has important implications for school administrators. One factor that contributed to ease in implementation was that the specified teacher behavior was readily observable (Stallings, 1974). By defining teacher behavior specifically, the person responsible for supervising implementation can readily assess whether the teacher is behaving appropriately and provide information to the teacher in the form of correction or reinforcement. On the other hand, if the teacher’s behavior is described in terms such as “observe and study individual children

## *THE EDUCATIONAL ESTABLISHMENT*

systematically and with insight and sensitivity” (Gilkeson, Smithberg, Bowman, & Rhine, 1981, p. 256), it is very difficult for an observer to ascertain whether this has occurred. It is difficult to know whether another person is “observing,” let alone with insight and sensitivity.

While proponents of less structured methods have argued that the Follow Through evaluation was “unfair” because the Direct Instruction and Behavior Analysis Models were more easily implemented, they seem to miss an important point. The ease with which a model could be implemented is part of the evaluation of that model. It would seem that one of the most important features of a model is that it be both easily implemented and effective. That was the intent behind the concept of planned variation in Follow Through. It was not enough to know that some sponsors could obtain some improvement in children’s performance in highly controlled laboratory settings. Methods were needed that could be implemented in individual schools.

In summary, although decisions about instructional methods are generally made at the district level, these decisions are seldom based on the criterion of student academic achievement. Administrators seem to be more motivated to adopt innovations that carry financial benefits than those associated with academic achievement. When innovations are introduced, they are likely to be similar to existing, and often ineffective, practices. In the event that an effective method were adopted, organizational problems preclude its effective implementation. The failure to arrange adequate training and supervision may be related to the failure to view teaching as a technology.

### **Teachers**

#### **Teacher Behavior and Achievement**

Observational studies of Follow Through classrooms have examined the relation between teaching behavior and student achievement. For example, Stallings (1973) found “that the children of . . . sponsors who emphasize behavior modification processes perform better on tests related to traditional academic skills than children of other sponsors” (p. 265). More specifically, Stallings (1975) determined that higher achievement scores were obtained in classrooms in which the teacher provided information, asked questions about the information, required the children to respond, and provided immediate feedback either by prompting the correct response or by praising a correct answer. Stallings concluded that highly controlled classroom environments in which teachers use systematic instruction and a high rate of positive reinforcement contributed to higher achievement scores. In short, the data obtained from Project Follow Through provide empirical evidence that student achievement is related to the behavior of the teacher. It seems apparent that the use of direct teaching methods will result in increased achievement in basic academic skills. However, it is generally accepted that these methods are not widely used by classroom teachers.

One reason that teachers have not sought out or used these methods is that they fail to recognize that their current methods are ineffective. One factor contributing to the tendency to accept current methods as adequate is that spurious evidence of the effectiveness of a given method may be provided by the fact that inevitably some students will learn when that method is used. A teacher receives partial reinforcement for

## PROJECT FOLLOW THROUGH

using personally preferred teaching methods when some children learn because of (or in spite of) the methods used. Understandably, when children succeed, the teacher rarely attributes the learning to factors beyond his or her control. The “proof” of the adequacy of the method is the success of selected students. Failure is attributed to deficiencies in the student, not to deficiencies in the teacher or the method.

In short, the teacher may contend that he or she has taught every student, and differences in learning must be attributed to differences in students. In fact, whether the teacher’s behavior may be called “teaching” depends on how one defines teaching. Teachers usually adopt definitions based on the lay vocabulary, which typically refer to teaching as “imparting knowledge.” Synonyms for “teach” include “guide” and “show.” In this view, the teacher leads the way or points out the path to knowledge and the pupil is expected to follow. Students who fall behind are explained as unmotivated, or, more commonly, unteachable. The teacher however, has fulfilled his or her responsibility by pointing the way.

The influence of this nonscientific vocabulary was evident in the descriptions of teaching behavior in the majority of Follow Through models. For example, in the Responsive Education Model, the teacher was to “guide [the child’s] discovery of solutions” (Emerick et al., 1973, p. 97). Among the teaching strategies listed for Bank Street teachers was to “impart knowledge” (Gilkeson et al., 1981, p. 256). Weikart and Banet (1975) described the role of the teacher in the Cognitively Oriented Curriculum Model as “facilitating children’s learning.”

The common feature of these descriptions of teaching activities is that they exclude any reference to the learner. Teaching is defined independently of the learner. By this definition, teaching refers to any and all activities in which the teacher engages (Johnston, 1973). The very definition of teaching then precludes the possibility that the teacher did not teach and places the responsibility for learning on the child.

An alternative definition of teaching has been proposed by adherents to a technology of teaching based on the science of behavior. From this perspective, “imparting knowledge,” or teaching, is defined as getting a person to behave in a particular way (Skinner, 1968). Learning is defined as a change in behavior and teaching involves arranging conditions under which those changes take place. The term “teaching” then is reserved for behavior that results in changes in the behavior of the learner. As Johnston (1973) pointed out, any distinction between “effective” and “ineffective” teaching methods becomes meaningless: “Methods which do not produce desirable changes in academic performance are simply not described as teaching . . .” (p. 7).

By this definition, the teacher must assume responsibility for the learning of each and every child. Teachers have resisted the notion that they be held accountable for the achievement of all students in part because they attribute student failure not to inadequate methods, but to variables that are beyond their control. For example, poverty may be considered the causal factor behind student failure. Keppel (1966) claimed that conditions of social and economic deprivation are often used as “alibis for failure to find effective ways to educate these children” (p. 39). Despite the fact that this reference is somewhat dated, the practice seems to be as prevalent today as when Keppel wrote. However, the data provided by the Follow Through experiment suggest that although teachers cannot

directly alleviate the conditions of poverty in which a child lives, it is nevertheless possible to alter the academic repertoires of the disadvantaged.

Student failure is often explained as due to individual or cultural differences. In particular students are assumed to differ in intelligence or abilities. When used as a defense for poor student performance, the individual differences argument takes the form that a particular child's skills lie in an area other than that sampled by the measurement instrument. Taken to its extreme, this argument is used to determine that students should be taught subjects for which they indicate some potential; if a child's strengths lie in areas other than academics then the child should be taught those subjects instead of academics. As a result, schools select students who learn with traditional instruction rather than look for methods to teach all students (Keller, 1978).

But the tendency to attribute student failure to the student may be due in part to a teacher's exasperation at being unable to produce learning. The teacher may in fact change her instructional approach when confronted with students who do not seem to be benefiting from instruction. The teacher's repertoire, however, is unlikely to include the skills necessary for teaching such difficult-to-teach children.

### **Inadequate Training and Supervision**

While it is commonly assumed that teachers have acquired the skills necessary to teach their students, in reality teachers are woefully unprepared to teach children who do not readily learn with traditional teaching methods. It has already been established that teachers are unlikely to receive adequate training in effective methods of instruction during their own professional preparation. Because the instructional methods a teacher uses are most likely to be those taught during his or her own training (Bijou, 1977a), it may be assumed that the majority of teachers enter the classroom with deficient teaching skills. In short, teachers have not been taught to teach effectively (Skinner, 1968; 1984). Teachers themselves seem to agree with this statement. Elementary school teachers have expressed the opinion that their experiences in college did not prepare them to teach effectively (Dornbusch & Scott, 1975).

Yet, those who propose that teachers be held accountable for the achievement of their students have failed to take this into consideration. Proposals for accountability have frequently been punitive. Such proposals seem to be based on the assumption that teachers, provided the proper motivation, would behave in ways that would result in improved performance of their students. But if teachers do not have the skills necessary for effective teaching, they will not be capable of teaching all students regardless of the consequences. Applying punitive consequences to organisms whose repertoires do not contain prerequisite behavior is not effective (Ulrich, Stachnik, & Mabry, 1974). If teachers do not have the skills necessary to teach every child, proposals that would institute social and economic sanctions are likely only to generate emotional behavior. It should be noted that this is analogous to the practice of invoking punitive consequences for children who do not acquire academic skills. These punishers take the form of failing grades, retention, and ridicule in the short run; the long term effect of growing up illiterate in today's society has more devastating and far-reaching consequences. In any event, it should not be surprising that teachers actively oppose measures that would punish them for not having the necessary teaching skills.

## *PROJECT FOLLOW THROUGH*

It may be argued that arranging consequences for teaching based on student achievement, either reinforcing or punitive, would motivate teachers to seek out the necessary training. But it is unlikely that teachers would know what type of training to look for, and if they did, it is doubtful they could easily find it. Although teachers' salaries are partially based on number of college credits, thus providing motivation to take additional courses, these courses are unlikely to result in improved teaching skills. Unless dramatic changes take place in the way colleges of education train teachers, teachers are not likely to acquire effective teaching skills by returning to the same institutions that failed to train them adequately in the first place.

Where then might teachers receive this training? One possible source is in-service training provided by the school districts. But, as has already been discussed, this training usually takes the form of workshops consisting of lectures about methods, not actual training in the use of those methods. As Follow Through sponsors discovered, workshops of this nature are inadequate for generating the needed change in teaching behavior (Weikart & Banet, 1975). The factor that seemed to be related to the implementation of a particular Follow Through model was continued in-service training. Teachers generally are not provided with this type of training and supervision by their school districts.

Lacking the needed training and supervision, teachers may be ineffective at using an instructional approach and thus may eventually abandon the method. It is not unusual to find Direct Instruction materials unused in classroom closets, or to hear a teacher say "I tried behavior modification but it didn't work." When teachers are inadequately trained and supervised, it should not be surprising that they are unsuccessful at using complex teaching methods. And, being unsuccessful, it is not surprising that they discontinue using those methods in favor of personally preferred, albeit ineffective, methods. Inadequate training and supervision are not the only reasons that teachers do not use effective teaching methods. Teachers may reject effective methods because of contrasting professional philosophies.

### **Conflicts With Existing Views of Teaching**

Stallings and Kaskowitz (1974) described the behavior of teachers in Direct Instruction and Behavior Analysis Follow Through classrooms as "quite atypical of generally practiced classroom behavior" (p. 220). Despite the fact that it is precisely because these models differ from traditional classroom practices that they are effective, this departure from generally accepted views of teaching may result in resistance from teachers. For example, Stallings (1974) found a significant negative correlation between implementation and satisfaction among Follow Through teachers using Direct Instruction methods. That is, as degree of implementation increased, satisfaction with the method decreased. Stallings found this relationship curious because it suggests that the teachers implementing the model most faithfully were least satisfied with it and wanted to change it. One feature that teachers may object to is the high degree of structure.

For example, Cronin (1980) interviewed teachers who had been required to use Direct Instruction for a period of two years. At the end of that time the teachers agreed that the students had learned more than had been believed possible. Yet some teachers were opposed to the method because they disliked the degree of structure. Evidence such

as this suggests that student outcomes are not sufficiently reinforcing to compete with other sources of reinforcement. As Skinner (1968) pointed out, only when student's academic achievement becomes important will effective teaching methods be sought out or used. It is, however, unlikely that student achievement will become important until effective contingencies are arranged for producing those outcomes.

### **Faulty Contingencies**

While it may be said that existing contingencies are faulty, it should be clear that all contingencies are effective in generating or maintaining some behavior. Thus, contingencies may be considered to be "faulty" only with respect to some goal. Contingencies experienced by teachers are generally unlikely to result in the use of methods that have been most effective in producing academic achievement. In general, there are few incentives for teachers to use effective methods and many disincentives. Some of these contingencies will be discussed.

Professional rewards are generally not based on student achievement. As a rule, teachers work in relative isolation. School principals do not monitor closely either the behavior of the teacher or the performance of students. Consequently, teachers receive little feedback from their superiors in the form of praise, recognition, or correction. In general, principals are more likely to be interested in whether the teacher can maintain order than whether he or she can teach effectively. Classroom management is a necessary condition for effective instruction but it is not sufficient for producing learning. In fact, "child-centered" approaches such as the Open Education and Responsive Education Follow Through models may provide adequate management of informal activities, but have "remained at a primitive level in the design of means to achieve learning objectives" (Bereiter & Kurland, 1981, p. 20). When reinforcement is contingent on effective management, rather than achievement, teachers may elect to use methods that result in an orderly classroom without respect to the effectiveness of those methods on achievement.

Tangible rewards are not based on student performance either. Teachers are paid for being in the classroom regardless of what takes place there. Pay is not based on the outcome of instruction, that is, student achievement. Pay scales are based on factors such as credentials, number of college credits beyond the credential, seniority, etc. There is no differential rate of pay based on quality of teaching, as measured by student academic achievement.

Teachers have little incentive, then, to engage in the behavior necessary to teach *all* of the children in their classroom. It is undoubtedly the case that more work is involved in teaching all children. It is also undoubtedly the case that teaching educationally disadvantaged children requires a great deal of work, not because they are economically disadvantaged, but because a great deal must be taught in a short amount of time if deficits are to be remediated. When teachers are required to "teach" only some of their students, their job is considerably easier than would be the case if they were held accountable for teaching all of their students. In cases in which the use of effective methods has been required, teachers have sought to escape the situation. For example, Cronin (1980) reported that some teachers who had been required to use Direct Instruction methods requested transfers to middle-class schools where there were fewer "difficult to teach" children, and consequently, less work involved for them. When



## PROJECT FOLLOW THROUGH

administrators approve such transfers they reinforce the continued use of ineffective teaching methods.

Administrators may fail to support the use of effective methods in other ways. Assuming that teachers were motivated to look for and use effective teaching methods, they may run into opposition within the school. For example, if Direct Instruction materials have not been approved for purchase by local administrators, a teacher will be unable to obtain those materials. While it is possible to use Direct Instruction techniques with other curriculum materials, most commercially available materials are not structured or sequenced in a manner that will result in maximum benefit to the student. Even if appropriate materials can be obtained, teachers may be forbidden by their superiors to use them. Carnine (1984b) reported a case in which a principal had demanded that teachers discontinue using Direct Instruction, although he had never seen it in use and refused to observe the teachers use it! As unbelievable as this sounds, it is unlikely that it is an isolated instance.

An additional way in which administrators and principals reinforce the continued use of ineffective teaching methods is by excusing teachers from teaching every student. When a teacher encounters students that are especially difficult to teach common practice is to refer the student out of her classroom. Notice that this referral process assumes the fault is with the learner. If there happens to be a teacher in the school who is using effective methods and being successful in teaching these difficult children, it is likely that the child will be moved to that room. In other words, effective teaching is punished by loading that teacher's classroom with difficult-to-teach children, resulting in more and harder work for the same amount of pay. At the same time teachers who cannot teach difficult children have been reinforced for *not* teaching that child.

In summary, teachers work in an environment in which there are few rewards for effective teaching. Supervisors do not offer approval for good teaching or disapproval for failure to teach. Schools are generally not organized to recognize or reward exemplary teaching thus providing teachers with little motivation to look for or use effective teaching methods.

### **The Public**

#### **Uninformed About Research**

Millions of dollars in taxpayers' money was spent on the Follow Through experiment to find effective methods for teaching disadvantaged elementary school children. In return for this expenditure, the government promised to use the results to provide a more effective education for all children. However, the government has reneged on their promise. Knowledge gained from the experiment has not been used to change educational policies. In short, the taxpayers have received a very poor return on their investment. Yet, there has been no move by the public to demand that the knowledge gained from Project Follow Through be used to improve educational practices. Why have Follow Through parents not demanded that program policies be changed to include only effective teaching methods? Why do parents whose children are in Title I or other compensatory programs not demand that effective methods be used to

teach their children? And why has the general public not demanded that effective methods be used to teach the disadvantaged in this country?

The most logical answer to these questions is that the public is unaware that effective teaching methods exist. Although parents and the school boards that represent them seem to be the groups with the greatest potential to demand the use of effective teaching methods, these groups cannot serve as advocates of effective teaching methods until they become informed that effective methods are available (Carnine, 1984b).

The general public, not surprisingly, is unaware of the results of research investigating different teaching methods. For example, although no formal surveys have been conducted, it is doubtful that more than a relative handful of citizens are aware of the existence of Project Follow Through, much less the findings of the evaluation, despite the fact that it was the largest educational experiment ever conducted. The results of the national evaluation are largely inaccessible to the general public. The majority of publications concerning the Follow Through evaluation have been in professional journals where they are not likely to be encountered by the public. Even if the public had access to the reports, the results are complex and not easily interpreted even by researchers and professional educators.

A few writers (Feinberg, 1977; Kilpatrick, 1977; McGrath, 1981) have summarized the findings of the national evaluation in publications intended for the public. Such publications are likely to be read by only a very small percentage of the public and then by those who are perhaps least likely to be concerned with compensatory education. A great many Americans for whom the results of the Follow Through evaluation may be especially relevant will never contact that information because they themselves are victims of an ineffective public education. An estimated 60 million Americans are what Kozol (1985) has referred to as "illiterate in terms of U.S. print communications" (p. 10). A large percentage of illiterate Americans are also poor people, those whose children would typically qualify for compensatory education programs. Consequently, it may be assumed that articles published in *The Washington Post*, *The Boston Globe*, or *Time Magazine* cannot be read by many of those for whom they are perhaps most important. In short, it is highly improbable that the public at large is informed that a technology of teaching exists that has the promise of educating children who have typically been considered unteachable.

### **Misinformed by Educators**

Because parents themselves are uninformed about educational methods, they turn for help to those presumed to be experts. However, relying on school personnel for information about their children has resulted in the public being misinformed about teaching methods. Principals attribute academic failure to the child, the parents, or society (Carnine, 1984b). In other words, they place the blame on just about any factor except inadequate teaching methods. Consequently, parents are left to believe that no methods exist that are effective with their children.

It is not surprising, then, that Follow Through parents were concerned that the failure of their children to perform well on achievement measures would be interpreted as further evidence that disadvantaged children are not teachable. The outrage with which the parents responded to the evidence that their children were not learning was

## *PROJECT FOLLOW THROUGH*

misdirected. No doubt prompted by the views of their sponsors, parents took issue with the whole concept of experimentation. They argued that their children were learning, despite empirical evidence to the contrary. What they did not consider was that the reason their children performed poorly on achievement measures in Project Follow Through was precisely the same reason disadvantaged children have historically fared poorly on such measures, that is, because the methods used were ineffective in teaching the skills measured by such tests.

Experts have misinformed the public by ignoring the importance of teaching method and maintaining that comprehensive services will result in equal educational opportunities for disadvantaged children. The individuals who have been most active in assuring the existence of Follow Through have not demanded that the government use the results to improve the education of their children, but have lobbied for continuation of the Follow Through program in general. These individuals have unwittingly argued for the continued funding of ineffective teaching methods. Their primary interest, however, seemed to be not in the continuation of the approach offered by a particular sponsor as much as in the continued federal funding made available to the schools that served as Follow Through sites. In fact, parents were generally not aware of the differences among Follow Through models (Weikart & Banet, 1975). While parents showed little interest in the various instructional approaches, they were interested in maintaining the school district's relationship with the sponsor.

It should be remembered that Follow Through combined planned variation of instructional methods with social services such as medical and dental care. Follow Through's deletion from the federal budget would result in the loss of revenue and of the services it provided to the school district. While nutritional, dental, and medical services are undoubtedly of great benefit to disadvantaged children, provision of such services has long been confused with educating disadvantaged children. The results of the Follow Through evaluation indicate that these services do not lead to academic success. Children who received these services and were provided with inadequate teaching methods performed at about the level that they would have in the absence of such services. This is not to say that social services such as those provided by Follow Through are not important to the health and well being of children. It seems clear, however, that they are not sufficient to bring about changes in academic achievement. What is necessary to produce increased academic skills is the use of effective teaching methods.

### **Cultural Relativism**

It has been argued that the question the Follow Through experiment was designed to answer, "What works best?", was inappropriate because it must be qualified to ask what works best to teach what. Camilli (1980) argued that what should be taught and how is a philosophical issue and that behavior on the part of the government to narrow the available educational choices is unacceptable because it infringes upon the rights of citizens to decide what the goals of education should be. Similarly, Mosteller (1975) argued that the multiplicity of goals for American education makes it difficult to evaluate different educational practices.

The concept of cultural relativism has long provided educators with an excuse for not educating disadvantaged children. Because educators have historically failed to teach

academic skills to disadvantaged and minority children, they have proposed that the problem is not that these children are culturally deprived but rather that they are culturally different. Furthermore, these cultural differences are assumed to be “built-in” and thus not subject to change (John-Steiner & Smith, 1979).

Benjamin (1981) has speculated that minorities have opposed Direct Instruction because the program assumes that minority children have cultural deficits, rather than cultural differences. But the method does not judge cultures. It simply assumes that a certain proficiency in basic academic skills is necessary to function in today’s society. It is pointless for any cultural group to maintain that they will not be judged by the standards of the culture at large. The practices of any culture will be judged by their survival value (Skinner, 1953). Supporting the educational establishment for their failure to teach children the skills that are necessary for survival in American culture can hardly be considered functional.

More recently, minority groups have begun to reject the idea that basic standards constitute cultural bias. There seems to be a growing awareness by some members of these groups that the contention of cultural bias can serve to maintain notions of the superiority of some cultural groups (Lyons, 1980). To claim that achievement measures are culturally biased is tantamount to claiming that any test on which one performs poorly is biased against individuals who do not do well on the test. Academic achievement tests have been criticized as culturally biased. But the tests are not biased; they are designed to identify those who have and have not learned the skills that are measured by the test. The question that must then be asked is why have certain children not learned those skills, and the answer is because the proper contingencies have not been arranged. Parents must first acknowledge the importance of basic academic skills and then demand that educators use teaching methods that are effective in teaching those skills to all children. To deny the importance of such skills is to relieve the schools of their responsibility to teach them.

### **Failure to Recognize the Scope of the Problem**

Members of the general public are likely to speak out strongly against current educational practices only when they are made aware that the problem of educating the disadvantaged is every person’s problem. However, there is a tendency among the general public to see the problem of teaching the disadvantaged as one that does not concern them. But the failure of the educational establishment to provide all children with the skills needed to succeed has far reaching consequences not only for the disadvantaged themselves, but for the culture in general.

While Follow Through was aimed at discovering methods of teaching the disadvantaged, that population need not be defined by socio-economic status. If the disadvantaged are characterized by the failure to benefit from “traditional” instruction, then we may define anyone whom the educational establishment fails to teach as instructionally disadvantaged. This condition is certainly not restricted to the poor. American schools are in serious trouble. The National Commission on Excellence in Education (1983) considered the nation at risk. The statistics included in the Commission’s report are sobering. An estimated 23 million American adults are functionally illiterate. (Kozol, 1985, suggests 60 million is a more realistic figure. This number represents more than one third of the adult population in this country.)

## *PROJECT FOLLOW THROUGH*

Approximately 17 percent of all 12-year old children are functionally illiterate. The percentage approaches 40 percent among minority children. Achievement test scores of high school students are lower than they were 26 years ago. And the national scores on these tests would be much lower if they included data from those students who drop out at an ever increasing rate.

Millions of dollars are spent annually on remedial education for military recruits and adult literacy programs. Millions of dollars are spent to support those who cannot perform at standards that will enable them to attain employment. Millions of tax dollars are spent on insurance to pay for accidents caused by the inability of an individual to read warnings or instructions. In short, illiteracy costs all Americans millions of dollars annually, and this does not even begin to measure the personal suffering of those individuals to whom the educational system has denied the opportunity to function in a literate society or the threat posed to democracy when many citizens are unable to participate in the election of their representatives.

In summary, the problem of the failure of American education is everyone's problem. The public has recently begun to react to the problem by proposing measures of teacher competence, but so far, those measures generally amount to literacy tests. That many American teachers have been found to lack basic academic skills is perhaps the ultimate indictment of the educational system. The public needs to be aware that the problems of education will not be remedied simply by discharging illiterate teachers. Literacy may be essential, but the schools are overrun with teachers who are competent in basic skills and yet cannot teach. The public must demand that teachers be trained in, and use, effective teaching methods.

This is likely to occur only when there is widespread agreement among Americans that the goal of education is academic achievement. While there seems to be increasing support for this position, there are still those who maintain that the purpose of schools is to serve as institutions of social change. Perhaps all citizens should ask themselves the provocative question Stallings and Hentzell (1978) raised: "If schools are used primarily as agents of social change, then what is the institution that teaches our children to read and write and do arithmetic? If the schools graduate the illiterate and the innumerate, then how much social change can we expect, and in what direction?" (p. 12).

## *SUMMARY AND CONCLUSIONS*

## **PART VI: SUMMARY AND CONCLUSIONS**

### **Problems with Project Follow Through**

#### **As an Experiment**

Project Follow Through was not initially conceived as an educational experiment, but as a social service program. The decision was made to change Follow Through from a social service program to an experiment prior to Congressional approval. When Congress did authorize Project Follow Through, it was not as an experiment, but as the social service program originally requested by President Johnson. This made Follow Through subject to a number of policy constraints, including the provision of comprehensive social services. The inclusion of these services resulted in ambiguity about the purpose of Project Follow Through. In addition, the requirements of the authorizing legislation restricted the ability to design a tightly controlled experiment. Although the administrators of Project Follow Through made no attempt to obtain a waiver of the legislative requirements, they later appealed to those requirements as an explanation for weaknesses in the design of the Follow Through experiment.

It seems reasonably clear that the willingness of program administrators to accept these statutory limitations was motivated by goals unrelated to designing an educational experiment. Administrators anticipated that Follow Through would eventually be expanded into a full scale social service program, provided enough support could be generated for such expansion. While the dream of expansion was never realized, the effects of efforts to generate widespread support for the program were.

Specifically, administrators sought to represent all educational interests. Consequently, they did not limit either the number or nature of educational approaches. The models represented great diversity and it was difficult to identify dimensions along which they varied. Rather than address the issue of specificity of treatment, administrators chose to ignore it. The problems generated by refusal to limit either the number of sponsors or the focus of the study became apparent with the effort to evaluate the models.

#### **Evaluation**

Early evaluation efforts were characterized by the degree to which the evaluation was controlled by interest groups, rather than by the experimental question. When early evaluation results were not favorable, the evaluation was criticized. Decisions were made to withhold the results from the public and to restrict measurement to children's entry and exit from the program. The latter decision made it difficult to assess differences between models, despite the fact that identifying differences in effectiveness was the purpose of the experiment. In general, administrators and evaluators seemed to ignore the original question and concentrate on meeting the demands of sponsors and interest groups whose goals were other than academic achievement for children. As evaluators tried to address the myriad interests, the evaluation grew to unmanageable proportions.

In 1972, changes occurred both in program administration and evaluation contractor, and the focus of the evaluation was restricted to the original question of what works to teach disadvantaged children. However, the evaluation continued to be

## *SUMMARY AND CONCLUSIONS*

constrained by early design decisions such as non-random selection of sites or students and non-random assignment. Because disadvantaged children, particularly those who had previously been enrolled in Head Start preschool programs, were assigned to Follow Through classrooms it was difficult for evaluators to construct comparison classrooms that were equivalent in socio-economic status and preschool experience. In order to control for these initial differences between treatment and comparison groups, evaluators found it necessary to conduct statistical manipulations on the data. This added to the complexity of the evaluation of Follow Through.

### **Summary**

One striking observation emerging out of the history of Project Follow Through is that none of the parties contributing to this history had a genuine empirical curiosity about which methods were most effective. They were not interested in learning whether any method was better than others in general or in any particular aspect. This was especially clear during the evaluation process when the parties involved conspired against the evaluation. Every effort was made to prevent an evaluation from being conducted that would reveal differences between models. Only after McDaniels assumed responsibility for the evaluation was the experimental question addressed.

The administrators of Follow Through naively expected all models to be equally effective. They expected a “slam bang” effect, predicting that all of the models would be tremendously effective compared to non-Follow Through classrooms. This suggests that those involved in Follow Through expected improvement in education to occur as a function of factors common to all models and absent from traditional classrooms. In particular, it can be assumed that they expected learning would be facilitated by money and comprehensive social services.

What this reveals is that Follow Through administrators, their advisors and consultants, evaluators, and in many cases the sponsors themselves, did not look at teaching method as a technology that has degrees of potential quality. This position derives from the premise that learning is a function of the learner, not an outcome of instruction. In other words, it is assumed that the critical mechanism for change is in the student, rather than in the student-teacher interaction. Therefore, administrators, educators, and evaluators failed to look at the method of instruction as a dimension of qualitative variation. They failed to see that the function of educational research is to determine what types of student-teacher interactions, or methods, result in learning.

Conducting Follow Through as an experiment, then, was not motivated by any true interest in obtaining knowledge about instructional methods, but by the desire to keep the program alive. Nevertheless, administrators did go to a great deal of trouble to introduce systematic variation in teaching approaches into classrooms throughout the country. Regardless of their motivation, an experiment did take place that has provided invaluable information about effective teaching methods.

Despite the design flaws noted, there were clear and consistent differences between models. The model with a preponderance of positive effects was the Direct Instruction model sponsored by the University of Oregon. The Behavior Analysis model ranked second in terms of percentage of positive effects for all measures. The Direct Instruction model demonstrated outcomes on measures of basic skills, cognitive skills,



## PROJECT FOLLOW THROUGH

and affective development that were significantly better than that of comparison classrooms and of other Follow Through models. The results of the evaluation did answer the question of what works best to teach disadvantaged children. In addition, the results of the national evaluation have been substantiated by secondary analyses of the data. The general conclusion that is stated consistently in every major evaluation of the Follow Through data is that the highest mean scores on achievement measures were realized by students enrolled in Direct Instruction and Behavior Analysis models.

### **Impact on Education**

#### **What Was Done**

The government failed to take action on the findings of their own study. Rather than endorsing only those models that were successful in increasing achievement of disadvantaged children, government officials proclaimed the program successful as a whole. All models continued to be funded by the federal government, including those models that had actually resulted in negative effects on achievement. In other words, the federal government used tax money to provide children with an education that results in greater academic deficits than would be expected if these children attended school in traditional classroom. This practice has continued despite the fact that the sponsors of those models have been able to provide no evidence of any offsetting advantages.

Officials in the Office of Education virtually ignored the findings. Plans were developed for continued research that would focus not on effective methods of instruction, but on models that could be easily implemented. OE seemed unaware of the fact that valuable information about the conditions necessary for implementing a model successfully had been provided by observational studies of Follow Through classrooms. OE officials ignored the finding that student achievement is related to instructional method. Rather than advertise this finding to the public, they buried the finding in a welter of bureaucratic qualifications.

Perhaps even more disturbing is the fact that the findings have been disregarded by the educational establishment. More than ten years following the publication of the Follow Through findings, there is little evidence that the results have altered educational practices in American classrooms. The only evidence that educators are even aware that the experiment occurred generally amounts to efforts to dispute the finding that child-centered or open education models are not effective alternatives to educating children. After examining the various facets of the educational establishment, it becomes clear that there has been little or no influence by the largest, most expensive, and most systematic effort in history to examine the effectiveness of various educational methods. If empirical evidence of the existence of effective instructional methods is not sufficient to prompt the use of those methods by the educational community, then we must look for additional or alternative sources of control.

#### **Factors That Influence the Educational Establishment**

*Policy Makers.* Policy makers are generally concerned with funding for federal programs, and funding is determined largely by support. The position federal officials adopt with respect to teaching methods is thus most likely to be congruent with the

## SUMMARY AND CONCLUSIONS

position of the majority. Because the Direct Instruction and Behavior Analysis models represent a minority view in education, it was not entirely unexpected that policy makers failed to take a strong position in support of the Follow Through results.

The administrators of federally funded programs are just that—administrators. Although some may have formal training in areas of education, they rely on input from professionals when developing programs. The influence of stakeholders in traditional educational practices can be seen throughout the history of the design, conduct, evaluation, and interpretation of Follow Through. Planning committees, advisory boards, and task forces were composed of representatives of universities and research centers. These professional educators generally represent philosophies that the Follow Through results suggest are not conducive to the development of effective teaching methods. For example, the chairman of the Follow Through National Advisory Committee was the Dean of the Bank Street College of Education whose model was ineffective in improving academic achievement or affective measures. The point is that those professionals who have the greatest influence on educational policy are likely to influence it in a direction that will not lead to improved education.

*Colleges of Education.* Professors and researchers who populate colleges of education have been particularly active in attempting to refute the findings of the Follow Through evaluation. One reason for this effort is that the most successful Follow Through models were derived from a philosophical orientation that is discordant with traditional philosophical views. While these effective methods may be rejected purely on philosophical grounds, there are more pragmatic reasons for objecting to any change in current practices.

Professional behavior consistent with traditional philosophies has been reinforced in the past. If contingencies were to change so that behavior consonant with a behavioral philosophy was reinforced, many college professors would be left with repertoires that would not enable them to be successful in attaining professional reinforcements. Thus, the reaction against research findings such as those of Follow Through may be an emotional reaction to the threat of loss of reinforcement. In addition, even if college professors were motivated to develop the necessary repertoire, few opportunities exist for acquiring such skills.

Despite the fact that the primary responsibility of college professors is teacher training, most of what takes place in colleges of education cannot be considered training in any real sense of the word. Instruction usually takes the form of lectures about learning and proceeds from the assumption that knowing about learning results in knowing how to teach. The Follow Through sponsors learned that lectures were ineffective in changing teaching behavior. Because their training consists primarily of instruction in this ineffective method, teachers enter their profession with inadequate skills.

*Teachers.* Observational studies of Follow Through classrooms suggested that exposure to the natural contingencies of the classroom does not generate effective teaching repertoires. The behavior of teachers in traditional non-Follow Through classrooms more closely resembled the behavior of teachers in the Follow Through models that were unable to improve academic achievement scores than that of teachers in the Behavior Analysis or Direct Instruction models. In short, teachers enter the classroom with ineffective skills and are unlikely to acquire these skills without specialized training.

## PROJECT FOLLOW THROUGH

They are, however, unlikely to seek out such training because in general professional rewards are not contingent on effective teaching.

Furthermore, many practitioners may not even be aware that methods exist that would enable them to be more effective. Even if they did know there was a better way to teach, how would they acquire the necessary skills? Those whom teachers depend on for training, college professors, have defective repertoires themselves. Project Follow Through has demonstrated that it is possible to retrain teachers. However, the circumstances under which these changes occurred were quite different from those typically found in schools.

*School Districts.* The primary goal of local administrators is the effective functioning of the system. Advocating anything other than prevalent practice would be disruptive, and administrators are unlikely to do so without some source of new motivation. This usually takes the form of outside funding. The funding provided to Follow Through sites undoubtedly influenced the decision of local districts to participate in the program. Follow Through differed significantly from other federally funded programs in that, in exchange for funding, the actual instructional practices were monitored. It was this system of monitoring and supervision that resulted in a high degree of conformity to different instructional models. Thus, while funding may provide the incentive to adopt an innovation, it may not be sufficient to ensure its implementation.

In Project Follow Through, monitoring and feedback systems were developed by model sponsors to maintain accurate implementation of the method of instruction. However, schools are generally not organized to provide this level of supervision. Although the organization of school systems appears to be hierarchical, in reality, teachers, principals, and administrators function independently. Principals and administrators have neither the motivation or the skills to function as instructional leaders. Instructional method is poorly controlled and typically left to the discretion of the teacher, whom we have already discovered is unlikely to have effective methods available to him or her.

*Publishers.* Much of what a teacher does instructionally is determined by the materials used. Educational materials are generally not constructed in a manner that is consistent with the sophisticated programming principles that characterize the instructional materials of the effective Direct Instruction model. The designers of instructional materials may not be expected to readily master the skills required to develop effective materials, and the probability that they will be called upon by publishers to do so is remote. The publishing industry does not typically act to initiate the development of unusual instructional materials, but reacts to the demands of the educational establishment. Publishing companies obtain information about current trends in education from their own consultants and advisors, education professors. The professors are in a good position to predict what educators will buy because their teaching regulates educational practices.

*The Public.* While the public is not typically considered part of the educational industry per se, it is included in this discussion because it supports education. What the public has supported has been the continued neglect of effective methods of instruction. Of course, the public has not done this knowingly; it is generally unaware that options exist.

## SUMMARY AND CONCLUSIONS

Parents and others rely on professional educators for information about education. They have been led to believe that the educational system is using the best methods available, and they have accepted that the failure of a great many students to learn is due to some deficit in the children. Members of the general public have no way of knowing that children are responding in ways that are perfectly compatible with the type of instruction that they are receiving.

### **Solving The Instructional Problem**

#### ***What Is Known***

Suggestions about how to address the problems of education have included changing the content of instruction, raising educational standards, increasing the amount of instructional time, increasing pay for teachers, and a long list of other “solutions” that would change just about every structural and functional aspect of education except *how children are taught*.

The major problem in education is that in thousands of classrooms throughout the country, children are being exposed to instructional methods that do not result in their learning basic academic skills that are necessary for survival in this society. In many classrooms, teaching methods are used that the Follow Through evaluation indicated result in greater deficits than might ordinarily be expected to occur. In some instances, the use of these methods is supported by funding from the federal government.

Educational researchers contend that they have been unsuccessful in producing a useful technology for teachers. While it may be the case that educators have failed to develop this technology, the technology does exist. The results of Project Follow Through indicate that a viable technology for teaching even the most difficult to teach children has been identified. The fact that a technology is available, however, does not ensure that it will be used. If it did, we would already have observed changes in educational practices and the academic performance of our children. The educational establishment has failed to use that which has the greatest potential for improving education, effective teaching methods. What is known is that methods exist that can solve some of the major problems of American education. What is necessary is the systematic use of those methods by the educational community.

#### ***Interlocking contingencies***

*Relating the elements.* As has been discussed, the various elements of the educational system are influenced by many different factors. The contingencies of reinforcement that govern the behavior of teachers, publishers, school administrators, government officials, members of the public, and professors of education are effective in maintaining ineffective educational practices. To ensure the systematic use of effective teaching methods, it is necessary to change the behavior of the “educational establishment.” We have learned from the experimental analysis of behavior that in order to change behavior we must change the contingencies of reinforcement.

Thus, in order to change the way the educational establishment responds with respect to teaching methods, it is necessary to alter the contingencies to which it is susceptible. Unfortunately, Skinner (1971) has predicted that it is nearly impossible to change the educational establishment. As difficult as the task may be, however, that is

## PROJECT FOLLOW THROUGH

precisely what must happen if we are to improve education. While it may be next to impossible to change the educational establishment as an entity, it may be feasible to systematically alter the practices of each element in a coordinated manner. Eventually, all facets of the educational system must change, but they are unlikely to change simultaneously or quickly.

In deciding where the most advantageous place to initiate change might be, it is necessary to consider the manner in which the various components of the educational system interact with and influence one another. While it has been discussed that each element is influenced by different factors, the parts of the system also influence one another. In other words, the elements function in an interlocking system of contingencies. Any attempt to change the behavior of the educational system must be preceded by analysis, at least at a nominal level, that identifies the behavior to be changed as well as the conditions that must be altered in order to change that behavior in desirable ways.

*Policy Makers.* Policy decisions are generally decisions concerning the disbursement of funds. Government officials make decisions about what types of educational research and programs will receive funding. Program administrators within federal agencies make decisions about the actual design of federally funded programs. Decisions concerning educational policy are not, however, made without input from the educational community. Individuals recognized as leading experts in the field of education serve as policy advisors. Grant review boards are composed of representatives of the universities. In other words, the behavior of government agencies is largely controlled by the academic community.

Policy decisions are also influenced by public support. Because those ultimately responsible for the decision to create, continue, modify, or eliminate federal programs are elected officials, they are susceptible to influence by popular opinion. In the case of educational programs, it may be expected that public opinion will be similar to that of the academic community because that is the public's primary source of information about educational policy. Follow Through is a case in point. Professional educators argued that models that failed to have a significant impact on outcome measures had not failed, but that the really important outcomes were not, or could not be, measured. In turn, parents testified before Congress that they knew their children were learning, and blamed the lack of evidence to support this contention on the measures used.

In summary, the behavior of policy makers is controlled by support from the academic community both directly and indirectly (through public opinion). In order to bring about change in the behavior of policy makers, it will be necessary to alter the behavior of members of the academic community.

*School Districts.* The decentralized system of control of American education places most of the authority for decisions concerning educational practices at the state or local level. Project Follow Through demonstrated that schools can be changed in ways that result in academic achievement for children, and it provided some information about the conditions necessary for change to occur.

One instructionally related behavior of representatives of school districts is the selection and adoption of textbooks. Curriculum materials should reflect the most advanced principles of instructional design. Only educational materials that have been extensively field tested, evaluated, and proven effective when used in certain ways should

## SUMMARY AND CONCLUSIONS

be adopted for use in the classroom. Unfortunately, few materials of this sort currently exist. The increased availability of appropriate instructional materials depends on change in the behavior of the publishing industry.

School districts are also responsible for hiring teachers. They establish certain qualifications for teachers, which should include the skills to use effective teaching methods. However, few teachers currently have the necessary skills because they are trained in colleges of education that rely on a defective teacher training paradigm. Thus, the availability of skilled teachers depends on changes in the behavior of the academic community responsible for teacher preparation.

In order to maintain effective instructional practices, school districts must establish systems to monitor teaching behavior and student academic achievement. Administrators and principals are unlikely to monitor instructionally relevant behavior of teachers because they have been trained by colleges of education to view teaching not as a technology, but as a creative and intuitive activity whose mechanism for change lies in the student. From this perspective, teaching is not seen as a controllable activity, and the need for monitoring either methods or student outcomes is not apparent. Principals and administrators must be taught that student achievement is functionally related to teacher behavior. This fundamental change in the way method is viewed will happen on a wide spread scale only when that point of view is taught in the institutions in which principals and administrators are trained.

*Teachers.* The changes described as desirable in the above section all have the intended goal of ensuring that teachers engage in effective methods of teaching. Current contingencies of reinforcement do not differentiate between effective and ineffective teaching (or teaching and not teaching). Some proposals have been made to arrange contingencies that would punish ineffective teaching and reward exemplary teaching. Such contingencies alone, assuming they could be arranged, are unlikely to generate the necessary behavior. The overwhelming majority of teachers do not have the necessary skills to be effective.

The way to bring about the desired changes in teacher behavior is not to punish deficient repertoires or to establish contingencies of reinforcement that the teacher can not contact due to behavioral deficits. The contingencies must wait until the teacher has acquired the skills to teach, and the most efficient way to provide teachers with the necessary skills is to teach those skills directly.

One way this might occur is through in-service training of some sort provided by school districts. While the typical workshop format has been demonstrated to be ineffective, other conditions could be arranged for continuing teacher training. There are limitations to approaching the problem at this level, however. The effects of any effort to retrain teachers currently in the field are limited to those teachers. Fortunately for the point at hand, the average number of years that each teacher stays in the classroom is quite limited, and thousands of new teachers enter the classroom each year. It is not a very efficient use of the limited resources of school districts to engage in activities designed to remediate (in a limited manner) deficiencies created by teacher training institutions. Teachers must enter the profession with the skills to teach every child. This will be accomplished only by systematic changes in what colleges of education teach, and how.

## PROJECT FOLLOW THROUGH

*The Public.* It may be assumed that parents want their children to receive the best education possible. However, parents lack the means to evaluate instructional practices. In fact, parents are generally unaware that there are differences in teaching methods, much less that methods are qualitatively different. Assuming that all teaching methods are created equal, parents are vulnerable to explanations of academic failure that appeal to some deficit in the child, such as perceptual problems, motivational problems, attention deficits, dyslexia, and a host of other “explanations” that serve only to convey to parents the message that the school is without fault. Of course, principals and teachers are responding to the situation in a manner that is consistent with their own professional training. The mechanism for learning is assumed to be inside the child; any failure to learn must then be due to a faulty mechanism or child.

Recently, there seems to be an increase in public awareness that something is “not right” in education. Parents, and the groups that represent them, have initiated proposals for teacher competency. Such proposals rarely take into account where this competency would come from. Teachers will be competent only when they are properly trained. Thus, if competent teachers are desirable, changes in teacher training programs must occur.

*Publishers.* The argument was made previously that in order for school districts to adopt technologically advanced curriculum materials, publishers must produce such materials. Engineering a change in the types of instructional materials that are available commercially will require changing the contingencies that maintain current practices. These contingencies are economic. Therefore, to induce publishers to alter their products it will be necessary to make it more profitable for them to produce materials based on advanced instructional design principles than the materials they currently produce. This will involve changing the buying behavior of the consumers of educational materials.

To do this, it must first be determined what maintains current buying patterns. Publishers produce two primary categories of instructionally relevant materials: Those that are used by college professors to train teachers and those used by teachers to teach children. It is clear that what is purchased in the first category is determined by college professors. In fact, these books are most often written by professors of education. While these books about teaching are unlikely to lead to effective teacher training, they are likely to continue to generate sales for publishers. If any change in books used to train teachers is to be realized, the behavior of the academic community must first be changed.

There is still the issue of textbooks purchased by school districts and used by classroom teachers. Again, what is produced is a function of what sells. What are considered appropriate instructional materials depends on how learning is conceptualized. If learning is viewed as an interaction between teacher/stimuli and student, then materials take on a critical role. If learning is believed to originate within the student, instructional materials are of little concern. The position adopted by teachers and school district personnel, those who select textbooks, is most often the latter. It has already been stated that teachers, principals, and administrators acquire this viewpoint during their professional training. Thus, if efforts to alter the behavior of publishers are to be successful, the academic community that trains the consumers of educational materials will need to be changed.

## SUMMARY AND CONCLUSIONS

This discussion should suggest that publishers are also pervasively influenced by professors of education. They not only create the materials, they advise publishers on where the field is going and what to publish by their service on advisory boards and as consultants and editors. While their influence is constrained by the actual sales of suggested materials, it is nevertheless considerable. Finally, they are also a primary source of sales, especially of college curriculum texts.

### **Colleges of Education**

While it is not exhaustive, this analysis of the contingencies that tie together all elements of the educational system as they concern educational method clearly points to the pervasive influence of the academic community. Colleges of education are the training ground for teachers, principals, administrators, and the entire miscellany of professional personnel that populate the educational establishment. Instructionally relevant behavior of teachers and school personnel is largely a function of this training. Professors of education also influence educational policy by serving on advisory and planning committees for federal programs such as Project Follow Through. In addition, they serve on review boards for granting agencies, thus influencing the types of educational research and development programs that are funded. Furthermore, members of the academic community influence state and local text book selection, thereby determining what materials will be used in classrooms. They impart this information to publishing companies by serving as advisors and consultants. Consequently, materials are published that are consonant with the views of college professors. Finally, professors of education represent the educational establishment to the public. It is clear that if any broad changes are to occur in the educational system, intervention must proceed at the level of the academic community.

### **How to Proceed**

*A Behavior Modification Metaphor.* The task at hand may be conceived of as a mammoth behavior modification project. The goal has been identified as not only the systematic use of effective instructional methods by teachers but all of the necessary supporting actions by other elements of the educational system. An extremely brief analysis has been done of the contingencies maintaining current behavior. In order to change behavior, the principles of behavior change must be systematically employed at both the strategic as well as the individual level. Because the requisite behavior for the use and support of effective instruction is not in the "repertoire" of the educational establishment, the effort must "begin where the learner is" by considering how existing contingencies can be improved. In this process, "behavioral educators" must function as trainers or behavior change agents.

This is an overly simplistic perspective in a sense, and yet it is the only one that is defensible and that will be functional. It requires examining existing as well as improved antecedent and consequent events and the contingencies in which they participate at both a system and an individual level. Because only a relatively small group of educators is in a position to engineer the necessary changes, they will be slow. It is therefore important that this small force work in a concerted manner. Encouragement may come from the fact that more effective teaching will be highly reinforcing to all



## PROJECT FOLLOW THROUGH

components of the educational system and from the prediction that the rate of change (acceleration) will be exponential.

*Books.* The case was made previously that instructional materials exert considerable control over the behavior of the classroom teacher. The same may be said of professors of education. Textbooks influence what is taught in rather obvious ways. Therefore, in order to change what teacher trainees are taught, it is necessary to change the instructional materials used by their professors. One way to accomplish this is for behavioral educators to inundate the textbook market with books presenting this perspective. The more prolific behavioral educators are in writing books, the greater the probability of one being selected for use by instructors.

The objective is to shift control of teaching behavior from the textbooks that currently are adopted to new textbooks. It is important, therefore, that textbooks authored by behavioral educators do not constitute such an abrupt change that stimulus control is disrupted. Language may be a special problem. The language of textbooks must be appropriate for the intended audience. Traditional educators have relied on the lay vocabulary and technical language may be confusing and alienating. Where technical terms are necessary, they must be carefully explained in terms that the reader understands.

*Journal Publications.* Another potential source of control over the behavior of the academic community is professional journals. A necessary condition for developing control by such stimuli is bringing traditional educators in contact with them. This will not occur if behavioral educators restrict their publications to behavioral journals that assure them a sympathetic audience. Behavioral educators will need to publish in the journals of the educational community.

In doing so, behavioral educators should be astute in determining the interests of traditional educators. If educators are interested in teaching creativity, for example, behavioral educators must conduct and publish research on teaching creative behavior. Where traditional educators have failed to be successful in finding an effective technology, behavioral educators must demonstrate to them the power of the available technology.

Journal publications should not be limited to research, but should include review and discussion articles as well. There are tremendous discrepancies between what educational journals present as the current state of instructional technology and what behavioral educators know is available. They must make this knowledge available to a broader audience by presenting the evidence of the technology and its accomplishments in journals that are widely circulated among the academic community.

Other articles should be aimed at educating the educational community about educationally-relevant philosophical issues that seem to generate unproductive emotional responses. These articles may stimulate discussion among colleagues and doctoral students. This latter group should be considered an especially important audience because they will be the educational leaders of tomorrow and because they do not yet have the extensive histories of their professors and thus may be more amenable to change.

Writing general articles that address how to think about educational issues may not be sufficient. Simply exposing the learner to such material is not likely to produce learning. The learner must be given an opportunity to respond. Behavioral educators must

## SUMMARY AND CONCLUSIONS

monitor the responses of the educational community by reading their journals. When inappropriate responses occur they must not go unaddressed. Correction must be provided in the form of further publications.

*Professional Meetings.* Books and articles have the advantage of providing a permanent response product, but they lack the potential for immediate correction and reinforcement that is provided by face-to-face contact. Fortunately, behavioral educators need not choose one tactic over another, they can, and should, engage in multiple forms of behavior that are likely to influence the academic community. In order to increase interactions with traditional educators, it will be necessary to go to them. Attending professional meetings will put behavioral educators in direct contact with the leaders in traditional education.

Behavioral educators can offer workshops at educational conferences demonstrating behavioral and direct instructional strategies and their results. Symposia can be arranged in which issues common to traditional and behavioral educators will be addressed. Panel discussions can be organized in which issues that interest all educators are discussed. Interactions of this nature allow for immediate correction of misunderstanding and misinterpretation. Our technology has shown us that the most effective interactions are direct interactions between the behavior change agent and the learner. Behavioral educators will need to arrange conditions that increase opportunities for these interactions.

*Demonstrations of Behavioral Instruction.* An important requirement for teaching is to provide the learner with examples. Behavioral educators need to do more than tell their traditionally trained colleagues about the possibilities afforded by a technology of teaching; they must show them. One way to do so is by establishing model classrooms and schools. Those who own or direct such schools should encourage departments of education to place practicum students in their schools. This will result not only in providing prospective teachers with direct training of instructional skills, but will also increase contact with the professors of education who are their supervisors.

Model classrooms should be made highly visible, both to educators and to the public. The accomplishments of behavioral educators will not command attention if they are not publicized. It will be difficult for educators to ignore the technology that can provide a solution to the major educational problems if those who support education know a solution is available.

*Funding.* Grant money is a powerful reinforcer for members of the academic community. Resources for educational research and development are limited, and as the available resources decrease, educators may be expected to compete for remaining monies. Behavioral educators should be active in their efforts to attain grant money. The probability of grants being awarded to behavioral educators increases greatly with the number of grants submitted.

Obtaining grant money will have two positive effects. First, it will reduce the availability of funding for traditional research efforts that have been unproductive in solving educational problems. If over time a majority of funding resources were distributed to behavioral educators, the research behavior of professors of education would change in accordance with the contingency. Second, grant money will provide the resources necessary for behavioral educators to establish more demonstration schools

## PROJECT FOLLOW THROUGH

resulting in even more widespread awareness of what can be accomplished in education with an effective technology of teaching.

In addition, funding will allow behavioral educators to conduct further programmatic research and development of the technology. The Direct Instruction and Behavior Analysis Follow Through models may represent the best teaching methods currently available, but they do not represent the best that is possible from a natural science of behavior. With limited resources available, it may be possible to generate an interest in collaborative research, an excellent way to familiarize those trained in traditional research methods with the methodology of the science of behavior.

*Create Demand.* Skinner (1984) suggested that if demonstrating the effectiveness of behavioral teaching methods is not sufficient to initiate changes in education then some form of "pressure" may be needed. One source of pressure was alluded to above in the discussion of demonstration classrooms. Parents and the general public may be reticent to express their dissatisfaction with education because they have been led to believe that what is offered is the best that is generally available. Behavioral educators must be active in educating the public. Demand from the parents of individuals with mental disabilities has resulted in tremendous change in the care and treatment of that population. Education of individuals with mental retardation now routinely involves the use of behavioral technology. Legislation has been enacted that guarantees handicapped children an appropriate education. It is within the realm of possibility that a similar situation can occur in education, provided that the public is informed that the means exist for providing an appropriate education to all children.

In addition to the general public, present teachers should be made aware that better (more effective) instructional methods than they were taught are available. One way to accomplish this is for behavioral educators to present workshops in schools and at teacher conferences. While a one-day workshop may not alter much behavior in and of itself, it can make teachers aware that effective methods do exist. Many teachers pursue advanced degrees, and most return to college for additional credits to advance their pay scale. Knowing that these methods exist, they may pressure departments of education to teach them. Another tactic is to write to representatives of teachers unions and other teacher organizations to make them aware that the training teachers receive does not represent the best that is available. Teachers are consumers of the university system, they pay thousands of dollars to be trained, yet they are graduated without the skills to do their job effectively. They should be encouraged by whatever means possible to hold the colleges of education accountable. Pressure and the demonstration of effective methods may motivate college professors to adopt more effective methods.

### Conclusion

The above suggestions represent only a few possible tactics for engineering change in the way professors of education think about, talk about, study, and most importantly, teach instructional methods. Although it may be desirable, it is not necessary that professors of education or teachers become thoroughgoing radical behaviorists. The sponsors of the Direct Instruction model do not fall into this category, yet they developed the most effective instructional method that is currently available. They could not have done so, however, had they not looked at teaching as a technology and at learning as an

## *SUMMARY AND CONCLUSIONS*

orderly process. It is this view of learning that is critical to convey to the educational community. Educators must be taught that learning is a function of the student-teacher interaction, the instructional moment. They must learn that there are qualitative variations in those interactions and that the function of educational research is to determine what types of interactions, or methods, lead to the most change with the least resources.

Another qualification is needed. While reference has been made to the academic community or to colleges of education as entities, it is obvious that colleges and departments of education are composed of individuals. Behavior can occur and be changed only at the level of the individual. The task at hand may not seem so overwhelming when it is approached in this manner. Fortunately, it may be expected that the number of faculty and teachers who adopt this perspective will accelerate exponentially. Every individual college professor whose behavior is changed will result in more teachers being trained effectively and consequently in more children learning. In addition, each of those college professors may join in the effort. Eventually the behavior of enough individuals may have been changed to create a new zeitgeist.

Finally, a treatise such as this cannot end without considering the effects of effective instructional methods for the survival of the culture. Although it is difficult to use this standard to evaluate cultural practices formally, it is hard to see any reasonable argument against the position that the quality of a culture's methods of passing on its accumulated knowledge to each generation is important to its survival. As cultural survival grows more difficult in an increasingly complex world, the knowledge that must be passed on grows in size and complexity. Conducting this dissemination effectively thus grows more difficult. The critical knowledge of this country three hundred years ago certainly did not demand the instructional sophistication that is necessary today. There is considerable evidence that we are increasingly unsuccessful in our educational effort relative to increasingly demanding survival criteria. Unfortunately, one of the difficulties of the cultural survival standard for a practice is that it takes a very long time to apply clearly. Indeed, ineffective practices may last for generations if not centuries. When will it be too late?

*PROJECT FOLLOW THROUGH*

## REFERENCES

### REFERENCES

- Alkin, M. C. (1970). *A review of the evaluation of the Follow Through program*. Los Angeles: Center for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 041 642)
- Anderson, R. B., St. Pierre, R. G., Proper, E. C., & Stebbins, L. B. (1978). Pardon us, but what was the question again?: A response to the critique of the Follow Through evaluation. *Harvard Educational Review, 48*, 161-170.
- Becker, W. C. (1977). Teaching reading and language to the disadvantaged - what we have learned from field research. *Harvard Educational Review, 47*, 518-543.
- Becker, W. C. (1978). The national evaluation of Follow Through: Behavior-theory-based programs come out on top. *Education and Urban Society, 10*, (4), 431-458.
- Becker, W. C., & Carnine, D. W. (1981). Direct Instruction: A behavior theory model for comprehensive educational intervention with the disadvantaged. In S. W. Bijou & R. Ruiz (Eds.), *Behavior modification: Contributions to education*. (pp. 145-210). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Beers, D. C. (1976). Principles of implementation: The lessons of Follow Through. Paper presented at the American Educational Research Association annual meeting. (ERIC Document Reproduction Service No. ED 124 283)
- Bell, T. H. (1983). Letter to The Honorable Ike Andrews. In *Proceedings of the Hearing on Follow Through Amendments of 1983*. Washington, DC: U.S. Government Printing Office.
- Benjamin, R. (1981). *Making Schools Work*. New York: Continuum Press.
- Bereiter, C., & Kurland, M. (1981). A constructive look at Follow Through Results. *Interchange, 12*, 1-22.
- Berman, P., & McLaughlin, M. W. (1975). *Federal programs supporting educational change: Vol. IV. The findings in review*. Santa Monica, CA: Rand Corporation.
- Bijou, S. W. (1977a). Practical implications of an interactional model of child development. *Exceptional Children, 44*, 6-14.
- Bissell, J. S. (1973). Planned variation in Head Start and Follow Through. In J. C. Stanley (Ed.), *Compensatory education for children, ages 2 to 8*. (pp. 63-85). Baltimore: Johns Hopkins University Press.
- Bock, G., Stebbins, L. B., & Proper, E. C. (1977). *Education as experimentation: A planned variation model, Vol. IV-B, Effects of Follow Through models*. Cambridge, MA: Abt Associates.
- Bronfenbrenner, U. (1967). *Memorandum: Comments on Wolff and Stein study*. (ERIC Document Reproduction Service No. ED 015 029).
- Bushell, D. (1978). An engineering approach to the elementary classroom: The Behavior Analysis Follow-Through project. In A. C. Catania & T. A. Brigham (Eds.), *Handbook of Applied Behavior Analysis* (pp. 525-563). New York: Irvington Publishers.
- Camilli, G. A. (1980). *A reanalysis of the effect of Follow Through on cognitive and affective development*. Unpublished doctoral dissertation. University of Colorado.
- Campbell, D. T., & Baruch, R. F. (1975). Making the case for randomized assignment to treatments by considering the alternatives. In C. A. Bennett & A. A. Lumsdaine (Eds.), *Evaluation and experiment: Some critical issues in assessing social programs*. New York: Academic Press.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand-McNally.
- Carnine, D. W. (1983). Government discrimination against effective educational practices. In *Proceedings of the Subcommittee on Human Resources Hearing on Follow Through Amendments of 1983* (pp. 99-103). Washington, DC: U.S. Government Printing Office.
- Carnine, D. W. (1984a). The federal commitment to excellence: Do as I say, not as I do. *Educational Leadership, 4*, 87-88.
- Carnine, D. W. (1984b). *Barriers to student achievement and how to overcome these barriers*. Unpublished manuscript.
- Cline, M. G., Ames, N., Anderson, R., Bales, R., Ferb, T., Joshi, M., Kane, M., Larson, F., Park, D., Proper, E., Stebbins, L., & Stern, C. (1974). *Final report: Education as experimentation, Evaluation of the Follow Through planned variation model. Vol. I-A: Early effects of Follow Through and Vol. I-B: Monographs*. Cambridge, MA: Abt Associates. (ERIC Document Reproduction Service Nos. ED 094-890 and ED 094-891)

## PROJECT FOLLOW THROUGH

- Cline, M. G., Ames, N., Anderson, R., Conte, M., Ferb, T., Hatt, C., Joshi, M., Larson, J., Park, D., Proper, E., Stebbins, L., & Stern, C. (1975). *Final report: Education as experimentation, Evaluation of the Follow Through planned variation model. Vol. II-A: Two-year effects of Follow Through and Vol. II-B: Appendices and monographs*. Cambridge, MA: Abt Associates. (ERIC Document Reproduction Service Nos. ED 108-768 and ED 108-769)
- Cohen, D. K. (1975). The value of social experiments. In A. M. Rivlin & M. P. Timpane (Eds.), *Planned variation in education: Should we give up or try harder?* (pp. 147-172). Washington, DC: The Brookings Institution.
- Cohen, D. K., & Garet, M. S. (1975). Reforming educational policy with applied social research. *Harvard Educational Review*, 45, 17-43.
- Coopersmith, S. (1967). *The antecedents of self-esteem*. San Francisco: W. H. Freeman.
- Crandall, V. C., Katkowsky, W., & Crandall, V. J. (1965). Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. *Child Development*, 36, 91-109.
- Cronin, D. P. (1980). *Implementation study, Year 2: Instructional staff interviews*. Los Altos, CA: John A. Emrick & Assoc.
- Donovan, J. C. (1967). *The politics of poverty*. New York: Pegasus.
- Dornbusch, S. D., & Scott, W. R. (1975). *Evaluation and the exercise of authority*. San Francisco: Jossey-Bass.
- Education as experimentation: A planned variation model. Volumes I-IV*. (1977). Cambridge, MA: Abt Associates.
- Edwards, R. H. & Bridewell J. L. (1979). *Following through on Follow Through*. Paper presented at the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 178 611).
- Egbert, R. L. (1971). Unpublished manuscript.
- Egbert, R. (1973). *Planned variation in Follow Through*. Paper presented to the Brookings Institution Panel on Social Experimentation, Washington, DC.
- Egbert, R. L. (1974). *Planned variation in Follow Through*. Prepared statement presented to the House of Representatives hearings on the Extension of Certain Programs under the Economic Opportunity Act of 1964.
- Elmore, R. F. (1975). The design of the Follow Through experiment. In A. M. Rivlin & M. P. Timpane (Eds.), *Planned variation in education* (pp. 23-46). Washington, DC: The Brookings Institution.
- Elmore, R. F. (1976). *Follow Through: Decisionmaking in a large scale social experiment*. Unpublished doctoral dissertation, Harvard University.
- Emerick, J. A., Sorensen, P. H., & Stearns, M. S. (1973). *Interim evaluation of the national Follow Through program 1969-1971*. Menlo Park, CA: Stanford Research Institute (ERIC Document Reproduction Service No. ED 086 371).
- Engelmann, S., & Carnine, D. (1982). *Theory of instruction: Principles and applications*. New York: Irvington Publishers.
- Feinberg, L. (1977, July 20). "Basic" teaching methods more effective, study says. *The Washington Post*, pp. 1, 8.
- Gaynor, A. L., Barrows, I., & Klenke, W. (1980). *A systems dynamics model of implementation of an innovation*. Madison, WI: Wisconsin Research and Development Center for Individualized Schooling, University of Wisconsin.
- Gelfand, M. I. (1981). The War on Poverty. In R. A. Divine (Ed.), *Exploring the Johnson Years* (pp. 126-154). Austin, TX: University of Texas Press.
- Gersten, R. (1984). Follow Through revisited: Reflections on the site variability issue. *Educational Evaluation and Policy Analysis*, 6, 411-423.
- Gilkeson, E. C., Smithberg, L. M., Bowman, G. W., & Rhine, W. R. (1981). Bank Street Model: A developmental-interaction approach. In W. R. Rhine (Ed.), *Making schools more effective* (pp. 249-288). New York: Academic Press.
- Goodrich, R. L., & St. Pierre, R. G. (1979). *Opportunities for studying later effects of Follow Through*. Cambridge, MA: Abt Associates.
- Goodrich, R. L., & St. Pierre, R. G. (1980). *Opportunities for studying later effects of Follow Through: Executive summary*. Cambridge, MA: Abt Associates.

## REFERENCES

- Gordon, I. J. (1972). An instructional theory approach to the analysis of selected early childhood programs. In I. J. Gordon (Ed.), *Early childhood education* (pp. 203-228). Chicago: University of Chicago Press.
- Gorham, W. (1967). *Report of the Task Force on Child Development*. Washington, DC: U.S. Government Printing Office.
- Greer, R. D. (1982). Countercontrols for the American Educational Research Association. *The Behavior Analyst*, 5, 65-76.
- Greer, R. D., Alessi, G., Graf, S. A., & Lindsley, O. (1983). *Task force report for follow up on Follow Through*. Paper submitted to the Council of the Association for Behavior Analysis.
- Guttman, D., & Willner, B. (1976). *The shadow government*. New York: Pantheon Books.
- Haney, W. (1977a). The Follow Through Experiment: Summary of an analysis of major evaluation reports. *Curriculum Inquiry*, 7 (3), 227-257.
- Haney, W. (1977b). *The Follow Through planned variation experiment, vol. 5: A technical history of the national Follow Through evaluation*. Washington, DC: Office of Education, U.S. Department of Health, Education and Welfare.
- Hodges, W. L., Sheehan, R., & Carter, H. (1979). Educational intervention: The role of Follow Through sponsors. *Phi Delta Kappan*, 58, 666-669.
- House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. (1978). No simple answer: Critique of the Follow Through evaluation. *Harvard Educational Review*, 48 (2), 128-159.
- John-Steiner, V., & Smith, L. (1979). *What do we know about teaching and learning in urban schools? Vol. 8: The educational promise of cultural pluralism*. Paper presented to the urban education program, CEMREL, Inc.'s, national conference on urban education. (ERIC Document Reproduction Service No. ED 185 163).
- Johnston, J. M. (1973). . . . *To beat back the frontiers of ignorance*. Presented at the First National Conference on Behavior Research and Technology in Higher Education, Georgia State University, Atlanta.
- Keller, F. S. (1978). Instructional technology and educational reform: 1977. *The Behavior Analyst*, 1, 48-53.
- Keller, F. S. (1985). A chat with Fred Keller. *The Behavioral Educator*, 4, 2-6.
- Kennedy, M. M. (1978). Findings from the Follow Through planned variation study. *Educational Researcher*, 7 (6), 3-11.
- Keppel, F. (1966). *The necessary revolution in American education*. New York: Harper and Row.
- Kilpatrick, J. J. (1977, July 1). Basics better in education, Cambridge group finds. *The Boston Evening Globe*, p. 19.
- Kozol, J. (1985). *Illiterate America*. New York: New American Library.
- Kruee, G. K. (1973). *An organizational analysis of Project Follow Through, final report*. Evanston, IL: Northwestern University. (ERIC Document Reproduction Service No. 093 446).
- Levitan, S. (1969). The community action program: A strategy to fight poverty. *Annals of the American Academy of Political and Social Science*, 395, 63-75.
- Lovitt, T. C. (1977). *In spite of my resistance, I've learned from children*. Columbus, OH: Charles Merrill.
- Lyons, G. (1980). Why teachers can't teach. *Phi Delta Kappan*, 62, 108-112.
- Maccoby, E. E., & Zellner, M. (1970). *Experiments in primary education: Aspects of Follow Through*. New York: Harcourt, Brace, Jovanovich.
- McCandless, B. R., & Evans, E. D. (1973). *Children and youth: Psychological development*. Hinsdale, IL: Dryden Press.
- McDaniels, G. L. (1975). Evaluation problems in Follow Through. In A. M. Rivlin & M. P. Timpane (Eds.), *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.
- McDill, E. L., McDill, M. S., & Sprehe, J. T. (1969). *Strategies for success in compensatory education: An appraisal of evaluation research*. Baltimore: The Johns Hopkins Press.
- McGrath, E. (1981, Nov. 9). "Pricklies" vs. "Gooeys": Conflicting theories of learning in the wake of Head Start. *Time Magazine*, p. 107.
- Metropolitan Achievement Tests*. (1970). New York: Harcourt, Brace, & Jovanovich.
- Meyer, J. W., Scott, W. R., & Deal, T. E. (1979). *Institutional and technical sources of organizational structure explaining the structure of educational organizations*. Paper presented at the Human Service Organizations, Center for Advanced Study in the Behavioral Sciences, Stanford, CA.
- Morris, V. C., Crowson, R., Hurwitz, E., & Porter-Gehrie, C. (1981). *The urban principal: Discretionary decision-making in a large educational organization*. Chicago: University of Illinois.



## PROJECT FOLLOW THROUGH

- Mosteller, M. (1975). Comments. In A. M. Rivlin & M. P. Timpane (Eds.), *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- Novak, J. D. (1977). *A theory of education*. Ithaca, NY: Cornell University Press.
- Organisation for Economic Co-operation and Development. (1980). *Reviews of national policies for education: United States federal policies for education of the disadvantaged*. Paris: OECD Publications.
- Parker, R., & Day, M. (1972). Comparisons of preschool curricula. In R. Parker (Ed.), *The preschool in action* (pp. 466-508). Boston: Allyn & Bacon.
- Pincus, J. (1974). Incentives for innovation in the public schools. *Review of Educational Research*, 44, 113-144.
- Proper, E. C., & St. Pierre, R. G. (1980). *A search for potential new Follow Through approaches: Executive summary*. Cambridge, MA: Abt Associates. (ERIC Document Reproduction Service No. ED 187 809).
- Ramp, E. A. (1983). Prepared statement presented to the House of Representatives Committee on Education and Labor. In *Proceedings of the Hearing on Follow Through Amendments of 1983* (pp. 38-57). Washington, DC: U.S. Government Printing Office.
- Raven, J. C. (1956). *Coloured progressive matrices*. Dumfries, England: The Crichton Royal.
- Rhine, W. R. (1981). Follow Through: Perspectives and possibilities. In W. R. Rhine (Ed.), *Making schools more effective: New directions from Follow Through* (pp. 291-324). New York: Academic Press.
- Rhine, R. W. (1983). The role of psychologists in the national Follow Through Project. *American Psychologist*, 38, 288-297.
- Riley, G. R. (1978). *Follow Through program evaluation, 1976-1977*. Wichita, KS: Wichita P. S. Division of research, planning, and development services.
- Rivlin, A. M. (1971). *Systematic thinking for social action*. Washington, DC: The Brookings Institution.
- Rivlin, A. M., & Timpane, M. P. (Eds.). (1975). *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.
- Rossi, P. H., Freeman, H. E., & Wright, S. R. (1979). *Evaluation: A systematic approach*. Beverly Hills, CA: Sage Publications.
- Rossi, P. H., & Williams, W. (Eds.). (1972). *Evaluating social programs: Theory, practice, and politics*. New York: Seminar Press.
- Sidman, M. (1960). *Tactics of scientific research*. New York: Basic Books, Inc.
- Seitz, V., Apfel, N. A., & Efron, C. (1977). *Long term effects of early intervention: The New Haven project*. Paper presented at the American Association for the Advancement of Science, Denver, Colorado.
- Skinner, B. F. (1953). *Science and human behavior*. New York: The Free Press.
- Skinner, B. F. (1957). *Verbal Behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Skinner, B. F. (1963). Reflections on a decade of teaching machines. *Teachers College Record*, 65, 168-177.
- Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts.
- Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Alfred A. Knopf, Inc.
- Skinner, B. F. (1984). The shame of American education. *American Psychologist*, 39, 947-954.
- Stallings, J. A. (1973). *Follow Through program classroom observation evaluation 1971-72*. Menlo Park, CA: Stanford Research Institute. (ERIC Document Reproduction Service No. ED 085 100)
- Stallings, J. A. (1974). *An implementation study of seven Follow Through models for education*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Stallings, J. A. (1975). Implementation and child effects of teaching practices in Follow Through classrooms. *Monographs of the Society for Research in Child Development*, 40.
- Stallings, J. A., & Hentzell, S. W. (1978). What do we know about teaching and learning in urban schools? *Effective teaching and learning in urban schools. Vol. 10*. (ERIC Document Reproduction Service No. ED 185 165)
- Stallings, J. A., & Kaskowitz, D. H. (1974). *Follow Through classroom observation evaluation (1972-1973)*. Menlo Park, CA: Stanford Research Institute.
- Stebbins, L. B., Anderson, R., Baltzel, C., Brooks, C., Evans, P., Ferb, K., Ferb, T., Freyheit, P., Freidman, S., Hagerty, M., Jastrzab, J., Katzenstein, E., Keefe, K., Ladner, R., Mackiernan, Y., Mingolla, E., Nelson, B., Piper, C., Poppe, D., Proper, E., St. Pierre, R., & Watkins, M. (1976). *Education as*

## REFERENCES

- experimentation: A planned variation model. Volume III-A: Findings, Cohort II and interim findings, Cohort III and Volume III-B: Appendices.* Cambridge, MA: Abt Associates.
- Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., and Cerva, T. R. (1977). *Education as experimentation: A planned variation model. Volume IV-A: An evaluation of Follow Through.* Cambridge, MA: Abt Associates.
- Tomlinson, T. M. (1981). The troubled years: An interpretive analysis of public schooling since 1950. *Phi Delta Kappan*, 62, 373.
- Ulrich, R., Stachnik, T., & Mabry, J. (Eds.) (1974). *Control of Human Behavior. Volume 3: Behavior modification in education.* Glenview, IL: Scott, Foresman and Company.
- U.S. Congress, Senate. (1974, May 29). Hearings before a subcommittee of the committee on appropriations. U.S. selects 30 school districts to "Follow Through" Head Start. (1967, July 2). *The New York Times*.
- Weikart, D. P., & Banet, B. A. (1975). Model design problems in Follow Through. In A. M. Rivlin & M. P. Timpane (Eds.), *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.
- Wholey, J. S. (1979). *New directions for Follow Through.* Interdepartmental memorandum, U.S. Office of Education, Washington, DC.
- Williams, W. (1972). *Social Policy Research and Analysis: The Experience of the Federal Social Agencies.* New York: Elsevier.
- Williams, W., & Evans, J. W. (1972). The politics of evaluation: The case of Head Start. In P. H. Rossi & W. Williams (Eds.), *Evaluating social programs* (pp. 247-264). New York: Seminar Press.
- Wisler, C. E., Burns, G. P., & Iwamoto, D. (1978). Follow Through redux: A response to the critique by House, Glass, McLean, and Walker. *Harvard Educational Review*, 48, 171-185.
- Wolff, M., & Stein, A. (1966). *Six months later: A comparison of children who had Head Start, summer 1965, with their classmates.* New York: Ferkauf Graduate School of Education, Yeshiva University. (ERIC Document Reproduction Service No. ED 015 125)
- Wright, R. J. (1975). The affective and cognitive consequences of an open education elementary school. *American Education Research Journal*, 12, 449-468.
- Zigler, E., & Valentine, K. (1969). An idea whose time had come. In E. Zigler & K. Valentine (Eds.), *Project Head Start: A legacy of the War on Poverty.* New York: The Free Press.