

EFFECTIVE School Practices

VOLUME 15, NUMBER 1

PRICE \$5.00

WINTER 1996

FOCUS: WHAT WAS THAT PROJECT FOLLOW THROUGH?

FROM THE PRESIDENT

- Bob Dixon 1

FROM THE FIELD

- Letters 3

OVERVIEW

The Story Behind Project Follow Through

- Bonnie Grossen, Editor 4

RESEARCH

Excerpts from the Abt Reports: Descriptions of the Models and the Results

- Geoffrey Bock, Linda Stebbins, and Elizabeth C. Proper, Abt Associates 10

A Constructive Look at Follow Through Results

- Carl Bereiter, Ontario Institute for Studies in Education
- Midian Kurland, University of Illinois at Urbana-Champaign 17

Sponsor Findings from Project Follow Through

- Wes Becker and Zig Engelmann, University of Oregon 33

Project Follow Through and Beyond

- Gary Adams, Educational Achievement Systems, Seattle, Washington 43

PERSPECTIVES

Follow Through: Why Didn't We?

- Cathy Watkins, California State University, Stanislaus 57

Our Failure to Follow Through

- Billy Tashman, New York, *Newsday* 67

Philosophy of *Effective School Practices*

1. Teachers are responsible for student learning.
2. The curriculum is a critical variable for instructional effectiveness.
3. Effective teaching practices are identified by instructional research that compares the results of a new practice with the results of a viable alternative.
4. Experiments should not be conducted using an entire generation of Americans. The initial experimentation with a new practice should be small in scale and carefully controlled so that negative outcomes are minimized.
5. A powerful technology for teaching exists that is not being utilized in most American schools.

Effective School Practices is published quarterly by the Association for Direct Instruction. The mission of the Association for Direct Instruction, as stated in the by-laws, is to promote the improvement of educational methods.

The name *Direct Instruction* originated with the highly effective instructional model first developed by Zig Engelmann in Project Follow Through during President Johnson's Great Society legislation. Although the evaluation of Project Follow Through showed the Direct Instruction model to be far more effective than the other models on every identified outcome, education in America remained generally unchanged.

A few educators, impressed by the extraordinary results of the original Direct Instruction model and the programs that were developed as DI evolved, formed the Association for Direct Instruction in 1981.

Today, this organization is a vanguard in promoting school practices that have been validated as effective through the use of the scientific method in educational research.

The Association for Direct Instruction was incorporated in 1981 in the state of Oregon for educational purposes. ADI is a nonprofit, tax-exempt

corporation under Section 501(c)3 of the Internal Revenue Code and is a publicly supported organization as defined in Sections 170(b)(1)(A)(ii) and 509(a)(1). Donations are tax-deductible.

A copy or summary of the current financial statement, or annual report, and registration filed by ADI may be obtained by contacting: ADI, PO 10252, Eugene, OR 97440 (503-485-1293). ADI is registered with the state of Oregon, Department of Justice, #79-16751. Copyright © 1996 Association for Direct Instruction.

ANNUAL SUBSCRIPTION RATES: \$20 U. S.; \$25 (U. S. currency or equivalent) Canada; \$30 Europe; \$40 airmail to Europe. (ISSN 1068-7378).

Editor Bonnie Grossen
Assistant Editor Jocelyn Warren
Associate Editors for Research Jerry Silbert
..... Russell Gersten
..... Geoff Colvin
..... Ed Kameenui

Proofreading and Typing .. Jocelyn Warren
Layout Jerry Marr
Printing Springfield News

e-mail: BGrossen@oregon.uoregon.edu

From the President

Bob Dixon

The following is the first in a series of profiles of ADI employees

The name "Bryan" is probably familiar to anyone who has ever phoned ADI or the Engelmann-Becker Corporation, or who has attended an ADI conference. But few people know much about Bryan, and his roles with both ADI and Engelmann-Becker. I'll tell you a little about him

Bryan Wickman is a thirty-five-year-old who first worked for the Engelmann-Becker Corporation part-time, after school, as a shipping clerk. Upon graduation from high school in June of 1979, he began to work full-time for Engelmann-Becker. His first job title was "receptionist." In my memory of those days, however, it seemed to me that Bryan was more a combination of Boy Friday, gopher, and whipping boy. He did all those jobs no one else wanted to do, and when anything went wrong, we could all yell, "WHERE'S BRYAN!"

As a testimony to both Bryan's thick skin and his native intelligence, he advanced his skills and position at Engelmann-Becker through the years, into positions such as office manager and production manager for the many instructional programs produced at "The Corp." Today, Bryan's various roles at The Corp include a seat on the board of directors, the office of treasurer, office manager, and production coordinator. One of Bryan's critical jobs is doing computer-based layout and design work for new programs and revisions.

Throughout his career at Engelmann-Becker, Bryan has been "on loan" to the Association for Direct Instruction. He assisted in the 5th annual Eugene Conference in 1979 and began managing the conference the following year. When ADI formed in 1981, Bryan continued to manage the Eugene conference (and other ADI conferences). He has managed every Eugene Conference since, with a brief hiatus in 1993. Overall, he has run at least fifty conferences for ADI all around the country. At one time or another, he has worked on every aspect of ADI business: bookkeeping and finance, products, production of *Effective School Practices* (formerly *ADI News*), and membership.

Because of Bryan's vast range of experiences at both The Corp and with ADI, and because of his long term of service, he is often called upon as the Corporate Memory for both organizations.



Bryan Wickman

Bryan is a husband (to Trish) and father of son C.J. and daughter Kasey. He taught both to read at early ages, using Teach Your Child to Read in 100 Easy Lessons, and has been an active father in many other ways as well. He coaches his daughter's soccer team, teaches computer classes at his children's school, he's a parent representative and chairperson of the Site Council at that school, and he's the President-elect of the parent-teacher association.

At some point over the years, Bryan's work stopped being a job and became a career. He's personally and professionally dedicated to the causes of Engelmann-Becker Corporation and the Association for Direct Instruction. Both organizations are fortunate to have him. What's in Bryan's future? He's giving some thought to running for the Eugene school board, and he will undoubtedly play major roles in what appears to be the very bright futures of The Corp and ADI. But no matter to what heights Bryan arises, some of us—whenever anything goes wrong—will instinctively yell out, even in an empty room, WHERE'S BRYAN!"

Bob Dixon

FROM THE PRESIDENT

Join a local ADI chapter

The persons below are organizing local ADI chapters. They plan to form local support groups and to sponsor local workshops, discussion groups, and newsletters. Contact the person nearest you for more information on local chapters. If your name is not on the list and you would like to form a local chapter, contact ADI, PO 10252, Eugene, OR 97440 or call (503) 485-1293.

Carolyn Cittamlet
1422 S. 13th St.
Philadelphia, PA 19147
Fax: 215-551-9790

Susan Kandell
212 S. Woodhams St.
Plainwell, MI 49080-1753

Kathleen Schaefer
2668 Tareyton Cr.
Stoughton, WI 53589

Patti Clark
Phoenix Academy
11032 Oak St.
Omaha, NE 68144

Paul Koeltzow
10318 Fern Dale Rd.
Dallas, TX 75238
214-341-5373

Diana Morgan/Thaddeus Lott
Wesley Elementary
800 Dillard St
Houston, TX 77091

Ardena Harris
5309 Vineyard Lane
Flushing, MI 48433

Clark Walker
300 West 100700
Ftu Green UT 84632

Ken Traupman
248 Nutmeg St.
San Diego, CA 92103

Anna Mae Gazo
3027 Ellen Ct.
Marina, CA 93933

Cathy Watkins
1956 La Linda Ct.
Turlock, CA 95380
cwatkins@koko.csustan.edu

Ursula Garrett
PO Box 241, Apt 169
Kahuku, Hawaii 96731

Chuck Main
PO Box 8
Silverdale, WA 98303

Betty Williams
Dept. of Special Education
AD Box 25

Gonzaga University
Spokane, WA 99258

Babette Engel
343 Dungeness Meadows
Sequim, WA 98382

Helen Munson, Tricia Walsh-
Caughlan
1603 NW 41st Circle
Camas, WA 98607

Larry Chamberlain
1063 Stelly's X Rd.
Brentwood 1324,
Vosiao, BC

Chuck Baxter
1085 Taugh
Ithaca, NY 14850
(607) 272-7957

Angus Lloyd
47 Brooklyn Ave.
Toronto, Ontario M4M 2X4
(416) 465-0606 (voice)
(416) 564-3324 (fax)



WHILE YOU'RE SURFING THE NET...

Check out the New Web Page: of the National Center to Improve the Tools of Educators (<http://darkwing.uoregon.edu/~ncite/>). Find valuable documents, research syntheses and information on free math workshops.

We now have **TWO Email Lists**: one for discussion and announcements (effschprac), another for announcements only (adinews).

To subscribe to the discussion and announcements list, send the following message from your email account:

To: Mailserv@oregon.uoregon.edu

Message: Subscribe effschprac

(Don't add *Please* or any other words to your message. It will only cause errors. Mailserv is a computer, not a person. No one reads your subscription request.)

By subscribing to the EFFSCHPRAC list, you will be able participate in discussions of topics of interest to ADI members. You will automatically receive in your email box all messages that are sent to the list. You can also send your news and views out to the list subscribers, like this:

To: Effschprac@oregon.uoregon.edu

Subject: *Whatever describes your topic.*

Message: *Whatever you want to say.*

To subscribe to the announcements only list (adinews), send from your email account the following message:

To: majordomo@lists.uoregon.edu

Message: subscribe adinews

On this list, you will receive announcements only, such as news of upcoming TV specials on DI, announcements from employers seeking persons with DI teaching skills and from those with DI teaching skills seeking jobs, and other news flashes.

From the Field

After 7 years toiling on my Ph.D., trying to figure out why children are not learning and teachers are getting paid more, I think I have come up with the best way of teaching children. I put them in a room with a Classic Literary Work, Homer's Odyssey or Iliad and I say, "Read these words." And, amazingly, they begin to read through osmosis.

If the preceding paragraph sounds exactly how teachers expect children to learn, you are right. Actually, I have recently read Ziggy's book on academic child abuse and I viewed your 20th annual Eugene Conference on DI and am sold on the idea. It is possible that my teachers in primary school used these techniques. I remember learning this way, but today kids seem to learn through discovery...gimme a break.

To make a long story short, I would like to find out more about the opportunity to attend one of your training sessions and / or find a school where Direct Instruction is being implemented here in St. Petersburg, Florida.

Thank you and Committed to teaching excellence,
Jaime David Dudash

Carefully Taught (by the Whole Word / Whole Language Method)
Nathan Crow
Seattle, Washington

[sung to the tune of "Carefully Taught," by Rodgers and Hammerstein,
from "South Pacific"]

You've got to be taught to guess and to pray
To God that your memory will know what to say,
'Cause plain English is taught like Egyptian these days,
You've got to be carefully taught.

You've got to be taught that each word is unique,
And not cry while they hide all the patterns you seek,
In meaningless worksheets with instructions in Greek,
You've got to be carefully taught.

[chorus] You've got to be taught before it's too late,
Before you are six or seven or eight,
While professors tell parents it's "phonics" you hate,
You've got to be carefully taught.

You've got to be taught not to be at your ease,
By whole language teachers with master's degrees,
Who make every lesson an agonized tease,
You've got to be carefully taught.

You've got to be taught to pretend you can read
The Big Book she holds as she meets every need
Of the two kids in class who can actually read,
You've got to be carefully taught.

[chorus]

FROM THE FIELD

The Story Behind Project Follow Through

Bonnie Grossen, Editor

Project Follow Through (FT) remains today the world's largest educational experiment. It began in 1967 as part of President Johnson's ambitious War on Poverty and continued until the summer of 1995, having cost about a billion dollars. Over the first 10 years more than 22 sponsors worked with over 180 sites at a cost of over \$500 million in a massive effort to find ways to break the cycle of poverty through improved education.

The noble intent of the fledgling Department of Education (DOE) and the Office of Economic Opportunity was to break the cycle of poverty through better education. Poor academic performance was known to correlate directly with poverty. Poor education then led to less economic opportunity for those children when they became adults, thus ensuring poverty for the next generation. FT planned to evaluate whether the poorest schools in America, both economically and academically impoverished, could be brought up to a level comparable with mainstream America. The actual achievement of the children would be used to determine success.

The architects of various theories and approaches who believed their methods could alleviate the detrimental educational effects of poverty were invited to submit applications to become sponsors of their models. Once the slate of models was selected, parent groups of the targeted schools serving children of poverty could select from among these sponsors one that their school would commit to work with over a period of several years.

The DOE-approved models were developed by academics in education with the exception of one, the Direct Instruction model, which had been developed by an expert Illinois preschool teacher with no formal training in educational methods. The models developed by the academics were similar in many ways. These similarities were particularly apparent when juxtaposed with the model developed by the expert preschool teacher from Illinois. The models developed by the academics consisted largely of general statements of democratic ideals and the philosophies of famous figures, such as John Dewey and Jean Piaget. The expert preschool teacher's model was a set of lesson plans that he had designed in order to share his expertise with other teachers.

The preschool teacher, Zig Engelmann, had begun developing his model in 1963 as he taught his non-identical twin boys at home, while he was still

working for an advertising agency. From the time the boys had learned to count at age 3 until a year later, Zig had taught them multi-digit multiplication, addition of fractions with like and unlike denominators, and basic algebraic concepts using only 20 minutes a day.

Many parents may have dismissed such an accomplishment as the result of having brilliant children. Zig thought differently; he thought he might be able to accomplish the same results with any child, especially children of poverty. He thought that children of poverty did not learn any differently than his very young boys, whose cognitive growth he had accelerated by providing them with carefully engineered instruction, rather than waiting for them to learn through random experience.

Zig filmed his infant sons doing math problems and showed the home movie to Carl Bereiter at the University of Illinois, where Carl was leading a preschool project to accelerate the cognitive growth of disadvantaged young children. Nothing was working. After seeing Zig's film, he asked Zig if he could accomplish similar results with other children. Zig said "yes" and got a job working with him.

Carl Bereiter decided to leave Illinois to go to the Ontario Institute for Studies in education. The preschool project needed a director with faculty rank, a ranking that Zig did not have, in order to continue to receive funding on a grant from the Carnegie Foundation.

Wes Becker, a professor of psychology saved the preschool by joining it as a co-director. Wes had graduated as a hot shot clinical psychologist from Stanford, having completed the undergraduate and graduate programs in a record six years. Wes had then moved from the orientation of a developmentalist to much the opposite, that of a behaviorist. At the time Wes became familiar with Zig's work Wes was doing a demonstration project to show how behavioral principles apply to human subjects. Wes's demonstration was having difficulties because the instructional program for teaching reading was not working (Sullivan Programmed Phonics). One of Wes's graduate students, Charlotte Giovanetti, also worked with Zig in the preschool. She told Wes, "We know how to do that," and proceeded to develop a small group program for teaching sounds in the Sullivan sequence. It was successful and impressed Wes.

As chance would have it, about the same time that Zig and Carl's preschool program was looking for a new director, Wes heard Jean Osborn describe the Direct Instruction program used in the preschool at a symposium. Wes personally commented to Jean afterward how taken he was with the careful analysis (building skills on preskills, choice of examples, etc.). That night he was attacked by phone calls, strategically planned, requesting him to replace Carl Bereiter. The callers assured him it would take only a little bit of his time.

So Wes agreed to a partnership that then consumed his life. Only a few months after Wes became involved in the preschool project with Zig, Project FT began. Wes and Zig became the Engelmann-Becker team and joined Project FT under the sponsorship of the University of Illinois in 1967.

Zig began sharing his expertise with other teachers in the form of the Direct Instruction System for

Excerpts from the home movie of Zig working with his twin sons were shown at the 1994 Eugene conference and are included in the Conference '94 videotape available through ADI. The Conference '94 tape also includes footage of Zig working with economically disadvantaged preschool children and comments from those who were there in the early days of Zig's career and FT.

Teaching Arithmetic and Reading (DISTAR or Direct Instruction). His phenomenal success started getting attention. Other talented people began working with Zig. Bob Egbert, who for years was the National Director of Project FT, describes a scene from those early days in a letter he wrote to Zig for the 20th anniversary celebration:

The University of Kansas was having its first summer workshop for teachers. Don Bushell had invited Ziggy to do a demonstration lesson. My image of that occasion is still crystal clear. Ziggy was at the front of the large classroom when a half dozen five-year-old children were brought in. They were shy in front of the large audience and had to be encouraged to sit in the semi-circle in front of Ziggy. "How in the world," I thought, "will

this large, imposing man who has not been educated as a teacher cope with this impossible situation?" I need not have been concerned. Within three minutes the excited youngsters, now on the edge of their chairs, were calling out answers individually or in unison, as requested, to the most "difficult" of Ziggy's challenges and questions. By the end of the demonstration lesson, the children had learned the material that Ziggy taught; they also had learned that they were very smart. They knew this because they could answer all of the questions that Ziggy had assured them were too hard for them! (The full text of Bob Egbert's letter is in the Fall, 1994 issue of *Effective School Practices* on page 20-21.)

Problems began to develop immediately with the University of Illinois' sponsorship. Illinois allowed no discounts for the large volume printing of materials that were sent to the schools. Furthermore, Illinois would not allow a Direct Instruction teacher training program as part of its undergraduate elementary education program. Teachers learning Direct Instruction could not get college credit toward teacher certification. Wes and Zig began looking for a new sponsor. They sent letters to 13 universities that had publicized an interest in the needs of disadvantaged children, offering their one and a half million dollar per annum grant to a more friendly campus. Only two universities even responded, Temple University in Pennsylvania and the University of Oregon. Being more centrally located, Temple seemed more desirable. But then the faculty of two departments at Temple voted on the question of whether Temple should invite the DI model to join them. The faculty were unanimously opposed.

That left only the University of Oregon in tiny remote Eugene, hours of flying time from all the sites. Bob Mattson and Richard Schminke, Associate Deans of the College of Education, expressed the eagerness of the University to have the Engelmann-Becker model come to Oregon. The DI project staff took a vote on whether to move to Eugene. At this point Zig voted against the move. (He hates to travel.) But he was outvoted. As if on signal, Wes Becker, along with a number of his former students who had started working on the project (Doug Carnine was one of those students), and Zig Engelmann, along with a number of his co-teachers and co-developers, left their homes in Illinois and moved to Eugene, Oregon in 1970.

The Effects of FT

One of the most interesting aspects of FT that is rarely discussed in the technical reports is the way schools selected the models they would implement. The model a school adopted was not selected by teachers, administrators, or central office educators. Parents selected the model. Large assemblies were held where the sponsors of the various models pitched their model to groups of parents comprising a Parent Advisory Committee (PAC) for the school. Administrators were usually present at these meetings and tried to influence parents' decisions. Using this selection process, the Direct Instruction model was the most popular model among schools; DI was implemented in more sites during FT than any other model. Yet among educators, DI was the dark horse. Most educators' bets would undoubtedly have been placed on any of the models but the Direct Instruction model. The model developed by the Illinois preschool teacher who didn't even have a teaching credential, much less a Ph.D. in education, was not expected by many educators to amount to much, especially since it seemed largely to contradict most of the current thinking. All sponsors were eagerly looking forward to the results.

The U.S. Department of Ed hired two independent agencies to collect and evaluate the effects of the various models. The data were evaluated in two primary ways. Each participating school was to be compared with a matched nonparticipating school to see if there were improvements. In reality, it became difficult to find matching schools. Many of the comparison schools were not equivalent on pretest scores to the respective FT schools. These pretest differences were adjusted with covariance statistics. In addition, norm-referenced measures were used to determine if the participating schools had reached the goal of the 50th percentile. This represented a common standard for all schools. Prior scores had indicated that schools with economically disadvantaged students would normally be expected to achieve at only the 20th percentile, without special intervention. The 20th percentile was therefore used as the "expected level" in the evaluation of the results.

The preliminary annual reports of the results were a horrifying surprise to most sponsors. By 1974, when San Diego School District dropped the self-sponsored models they had been using with little success since 1968, the U.S. Department of Ed allowed San Diego only two choices—Direct Instruction or the Kansas Behavioral Analysis model. It was evident by this time that the only two models that were demonstrating any positive results were these two. The results of the evaluation were al-

ready moving into policy. This was not well-received by the many sponsors of models that were not successful.

Before the final report was even released, the Ford foundation arranged with Ernest House to do a third evaluation—a critique of the FT evaluation—to discredit the embarrassing results. The critique was published in the *Harvard Educational Review* and widely disseminated.

Ernest House describes the political context for this third evaluation as follows:

In view of the importance of the FT program and its potential impact on education, a program officer from the Ford Foundation asked Ernest House in the fall of 1976 whether a third-party review of the FT evaluation might be warranted. FT had already received considerable attention, and the findings of the evaluation could affect education for a long time to come. Although the sample was drawn from a nonrepresentative group of disadvantaged children, the findings would likely be generalized far beyond the group of children involved. Moreover, while the study had not yet been completed, the evaluation had generated considerable controversy, and most of the sponsors were quite unhappy with preliminary reports. Finally, the evaluation represented the culmination of years of federal policy, stretching back to the evaluation of Head Start. Would this evaluation entail the same difficulties and controversies as previous ones? Would there be lessons to be learned for the future?

For these reasons and after examining various documents and talking to major participants in the evaluation, House recommended that a third-party review would be advisable. If such a review could not settle the controversies, it could at least provide another perspective. The evaluation promised to be far too influential on the national scene not to be critically examined.

In January 1977 the Ford Foundation awarded a grant to the Center for Instructional Research and Curriculum Evaluation at the University of Illinois to conduct the study, with Ernest House named as project director. House then solicited names of people to serve on the panel from leading authorities in measurement, evaluation, and early-childhood education. The major selection criteria were that panel members have a national reputation in their fields and no significant

affiliation with FT. The panelists chosen by this procedure were Gene V. Glass of the University of Colorado, Leslie D. McLean of the Ontario Institute for Studies in Education, and Decker F. Walker of Stanford University. (p. 129, House, Glass, McLean, & Walker, 1978)

The main purpose of House et. al.'s critique seemed directed at preventing the FT evaluation results from influencing education policy. House implied that it was even inappropriate to ask "Which model works best?" as the FT evaluation had: "The ultimate question posed in the evaluation was 'Which model works best?' rather than such other questions as 'What makes the models work?' or 'How can one make the models work better?'" (p. 131, House, Glass, McLean, & Walker, 1978).

Glass wrote another report for the National Institute of Education (NIE), which convinced them not to disseminate the results of the FT evaluations they had paid 30 to 40 million dollars to have done. The following is an ERIC abstract of Glass's report to the NIE:

Two questions are addressed in this document: What is worth knowing about Project FT? And, How should the National Institute of Education (NIE) evaluate the FT program? Discussion of the first question focuses on findings of past FT evaluations, problems associated with the use of experimental design and statistics, and prospects for discovering new knowledge about the program. With respect to the second question, it is suggested that NIE should conduct evaluation emphasizing an ethnographic, principally descriptive case-study approach to enable informed choice by those involved in the program. The discussion is based on the following assumptions: (1) Past evaluations of FT have been quantitative, experimental approaches to deriving value judgments; (2) The deficiencies of quantitative, experimental evaluation approaches are so thorough and irreparable as to disqualify their use; (3) There are probably at most a half-dozen important approaches to teaching children, and these are already well-represented in existing FT models; and (4) The audience for FT evaluations is an audience of teachers to whom appeals to the need for accountability for public funds or the rationality of science are largely irrelevant. Appended to the discussion are Cronbach's 95 theses about the proper

roles, methods, and uses of evaluation. Theses running counter to a federal model of program evaluation are asterisked. (Eric Reproduction Service ED244738. Abstract of Glass, G. & Camilli, G., 1981, "FT" Evaluation, National Institute of Education, Washington, DC).

"The audience for FT evaluations is an audience of teachers to whom appeals to the need for accountability for public funds or the rationality of science are largely irrelevant."

ERIC abstract of

Gene V. Glass's critique

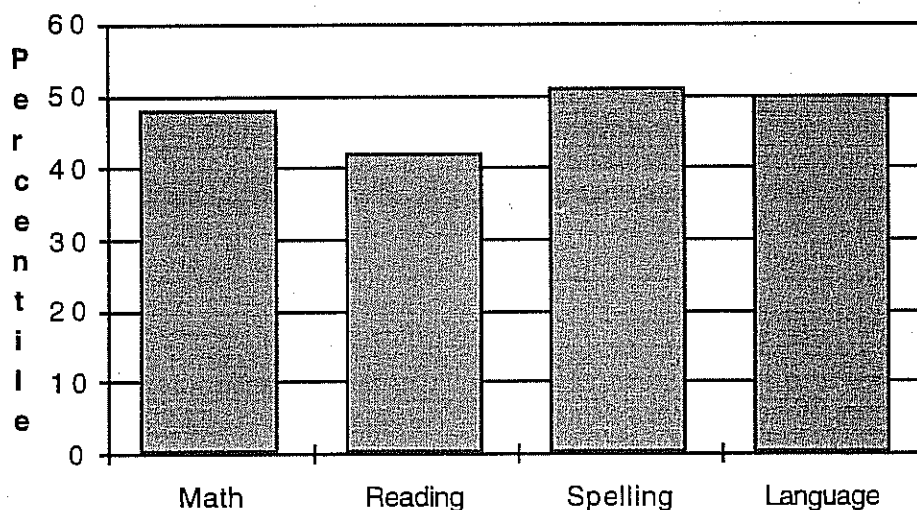
The final Abt report (Bock, Stebbins, & Proper, 1977) showed that the aggregate effects of all the models rendered FT to be a failure. FT was a failure because all of the models, except one, did not produce the desired results. (The Kansas Behavioral Analysis model also got positive results, but they were not as strong as the Direct Instruction model.) However, the FT Project did successfully identify what does work. The only model that brought children close to the 50th percentile in all subject areas was the Direct Instruction model.

These remarkable results were achieved by the Direct Instruction model in spite of the fact that Grand Rapids, MI was included in the analysis. The PAC in Grand Rapids had originally chosen to participate in FT using the Direct Instruction model. A new superintendent to the district later convinced the PAC to reject the model. The Direct Instruction sponsors subsequently withdrew from Grand Rapids; however, the US Office of Education continued to fund the site and continued to categorize it as Direct Instruction. It is probably not irrelevant that at this time Gerald Ford from Michigan was the U.S. President. In any case, because Grand Rapids had received FT funding throughout the evaluation period (1971-1976), they were included in the Abt analysis even though they had not implemented Direct Instruction for several years.

The most popular models were not only unable to demonstrate many positive effects; most of them produced a large number of negative effects. (See articles in this issue for details.)

After the House-Glass critiques were published, Bereiter and Kurland reviewed the FT data once again in 1981-2, responding in detail to each ques-

National Percentile Scores for the Direct Instruction Follow Through Model



Note: These schools achieved at the 20th percentile prior to Follow Through

tion and issue raised by House-Glass in a comprehensive and very readable report of the results. This review is reprinted in full on pages 17-32.

In spite of the counter arguments raised by Bereiter and Kurland and others, the House-Glass critique was successful. The results of Project FT were not used to shape education policy. Though much of the House and Glass critiques were based on a rejection of the use of experimental science in education, other critics, who did not reject experimental science, argued that the outcomes valued by the losing approaches had not been measured in the FT evaluation. Though some pleaded for more extensive evaluation studies of multiple outcomes, no further evaluation was funded. The following excerpts from Bob Egbert's letter to Zig provide his perspective on the evaluation:

No one who was not there during the early years of Head Start and FT can know how much your initiative, intellect and commitment contributed to the development of those programs. You simply shook off criticism and attempts at censorship and moved ahead, because you knew you were right and that what you were doing was important for kids.

Lest you think that censorship is too strong a word, let me remind you that many in the early education field did not want your program included in FT. As confirming evidence for my personal experience and memory I cite the Head Start consultant meeting held in, I think, September 1966, in which a group of consultants, by their shrill complaints, stopped

the full release of a Head Start Rainbow Series pamphlet which described an approach more direct than the approach favored by mainline early childhood educators—but one that was much less direct than the one you and Carl Bereiter were developing and using. The endorsement of Milton Akers for inclusion of "all" approaches in Head Start and FT Planned Variation made our task much easier.

Ziggy, despite what some critics have said, your program's educational achievement success through the third grade is thoroughly documented in the Abt reports. Your own followup studies have validated the program's longer term success. I am completely convinced that more extensive studies of multiple outcomes, which the Department of Education has been unwilling to fund, would provide a great deal more evidence for your program's success.

After the Abt report in 1977, there was no further independent evaluation of FT. However, the DOE did provide research funds to individual sponsors to do follow-up studies. The Becker and Engelmann article on pages 33-42 summarize the results of the follow-up studies by the Direct Instruction sponsors. Gary Adams' summary of the various reports of the results of FT provides a discussion of the reasons for the different reports and the consistencies and differences across them. This summary is excerpted from a chapter on Project FT research in a new book summarizing Direct Instruction research (Adams & Engelmann, *Direct Instruction Research, Educational Achievement Systems*).

FT and Public Policy Today

In responding to the critique by House et al., Wisler, Burns, & Iwamoto summarized the two important findings of Project FT:

With a few exceptions, the models assessed in the national FT evaluation did not overcome the educational disadvantages poor children have. The most notable exception was the Direct Instruction model sponsored by the University of Oregon.

Another lesson of FT is that educational innovations do not always work better than what they replace. Many might say that we do not need an experiment to prove that, but it needs to be mentioned because education has just come through a period in which the not-always-stated assumption was that any change was for the better. The result was a climate in which those responsible for the changes did not worry too much about the consequences. The FT evaluation and other recent evaluations should temper our expectations. (p. 179-181, Wisler, Burns, & Iwamoto, 1978).

The most expensive educational experiment in the world showed that change alone will not improve education. Yet change for the sake of change is the major theme of the current educational reform effort. Improving education requires more thought than simply making changes.

Perhaps the ultimate irony of the FT evaluation is that the critics advocated extreme caution in adopting any practice as policy in education; they judged the extensive evaluation of the FT Project inadequate. Yet 10 short years later, the models that achieved the worst results, even negative results, are the ones that are, in fact, becoming legislated policy in many states, under new names. Descriptions of each of the models evaluated in FT, excerpted from the Abt report, are included in this issue on pages 10-16. The Abt Associates ensured that these descriptions were carefully edited and approved by each of the participating sponsors, so they would accurately describe the important features of each of the models. Any reader familiar with current trendy practices that are becoming policy in many areas of North America, will easily recognize these practices in the descriptions of models evaluated in Project FT, perhaps under different names.

Curriculum organizations, in particular, are working to get these failed models adopted as public

policy. The National Association for the Education of Young Children (NAEYC), for example, advocates for legislative adoption of the failed Open Education model under the new name "developmentally appropriate practices." This model has been mandated in Kentucky, Oregon, and British Columbia. Oregon and British Columbia have since overturned these mandates. However, the NAEYC effort continues. Several curricular organizations advocate the language experience approach that was the Tucson Early Education Model in FT, under the new name "whole language."

That these curricular organizations can be so successful in influencing public policy, in spite of a national effort to reach world class standards and the results of scientific research as extensive as that in FT, is alarming. That the major source of scientific knowledge in education, the educational research program of the federal government, is in danger of being cut is alarming.

That the scientific knowledge we have about education needs to be better disseminated is clear. At the very least the models that failed, even to the point of producing lower levels of performance, should not be the educational models being adopted in public policy.

I, personally, would not advocate mandating Direct Instruction, even though it was the clear winner. I don't think that mandates work very well. But every educator in the country should know that in the history of education, no educational model has ever been documented to achieve such positive results with such consistency across so many variable sites as Direct Instruction. It never happened before FT, and it hasn't happened since. What Wes, Zig, and their associates accomplished in Project FT should be recognized as one of the most important educational accomplishments in history. Not enough people know this.

References

- Wisler, C., Burns, G.P., Jr., & Iwamoto, D. (1978). FT redux: A response to the critique by House, Glass, McLean, & Walker. *Harvard Educational Review*, 48(2), 171-185.
- House, E., Glass, G., McLean, L., & Walker, D. (1978). No simple answer: Critique of the FT evaluation. *Harvard Educational Review*, 48(2), 128-160.
- Bock, G., Stebbins, L., & Proper, E. (1977). *Education as experimentation: A planned variation model (Volume IV-A & B) Effects of follow through models*. Washington, D.C.: Abt Associates.

Excerpts from the Abt Reports: Descriptions of the Models and Summary of Results

Education as Experimentation: A Planned Variation Model

Geoffrey Bock, Linda Stebbins with Elizabeth C. Proper

Abt Associates

April 15, 1977

EDITOR'S NOTE: The following excerpts from the final evaluation reports of Project Follow Through include a description of each model and a summary of its results.

Volume IV-B:

Effects of Follow Through Models

The information [for the descriptions of the models] was taken from several sources including personal communication with the sponsors or their representatives...Each sponsor also had the opportunity to edit [the descriptions]. Many sponsors have expended considerable effort in rephrasing our materials to ensure their accuracy. We are grateful for their assistance and have tried to abide by their perceptions. (page 4)

Responsive Education Model Far West Laboratory for Educational Research and Development

The Model

The goals of the Responsive Education Model are for learners to develop problem solving abilities, healthy self-concepts, and culturally pluralistic attitudes and behaviors. Attainment of these goals and program objectives requires that the learning environment support productive child-centered learning and that the curriculum content and skills be relevant to the children's experiences outside the classroom.

The essence of this program has been described as follows:

The Responsive Program is based on beliefs in building a pluralistic society and in strengthening children as individuals. Instead of the

deficit view of compensatory education that focuses on deficiencies of low-income minority children, it adheres to a productive approach of enhancing the values of cultural differences and responding to the strengths of children as individuals...Schools should respond to children and their families rather than vice versa. (Judd and Wood, 1973)

The Responsive Education Model assumes that no single theory of learning can account for all the modes in which children learn; therefore, it seeks to provide a variety of learning experiences which build on the background, culture and lifestyle the child brings into the classroom. The child in a responsive learning environment engages in exploring, raising questions, planning, making choices and setting goals. The child discovers individual self-strengths, preferences, and liabilities. Each child develops a repertoire of abilities for building a broad and varied experiential base as well as self-confidence. The child interacts with all aspects of the educational environment, including other children. Whether individually, or within a group, the child may take on the role of leader, follower, or evaluator. These interactions can be curriculum oriented and may also involve personal and social issues. As the child grows through learning experiences, which address personal and social issues, inquiries are made into the nature of problem solving and the child takes greater responsibility for learning.

The teachers are integral and key contributors in

a responsive learning environment. They are skilled observers of the learners in a manner that supports and contributes to the objectives and principles of the Responsive Education Model. The teachers in this model establish an educational climate, develop a curriculum, and facilitate the learner's experiences.

The Responsive Education Model emphasizes the use of parents in meeting the program's objectives. Parents are encouraged to share in policy and curriculum decisions, to participate in the Parent Advisory Council (PAC), and to become involved in the classroom. This program provides specific training to help parents extend program objectives in the home. Parents are taught to use games and toys checked out from a toy library (maintained to provide parents with materials), to teach concepts contributing to program objectives. Parents also meet in workshops where they are taught to make learning aids. Through this training and through the volunteer classroom activities, parents have the opportunity to learn those types of adult-child interactions consistent with the objectives of the program.

The Responsive Education Model as

Realized in Follow Through

The Responsive Education Model is evaluated in eleven sites: Berkeley, CA; Buffalo, NY; Duluth, MN; Fresno, CA; Lebanon, NH; Salt Lake City, UT; St. Louis, MO; Tacoma, WA; Goldsboro, NC; Sumter, SC; and Owensboro, KY.

Tucson Early Education Model (TEEM)

Arizona Center for Early Education

The Model

The Tucson Early Education Model (TEEM) is based on the concept that each child has a unique growth pattern with individual rates and styles of learning. Based on M. Hughes' idea that formal learning should begin with the experiences young children bring to the classroom, and that the children's understanding of words and their meanings depends on the children's experiences, TEEM emphasizes a language-experience approach to cognitive development (Judd and Wood, 1973). The classroom is designed to support the use of language in relating experiences and learning how to learn.

Teachers work with children in groups of three to six. These groups are deliberately heterogeneous so that children will learn from peers. Interest centers are provided in the classroom to stimulate discovery and learning. Some classroom activities are selected and structured by the teacher; others are chosen by the children. Both types of activities are

based on student need and interest. Even in this open-ended context the learning experiences of the children are carefully structured through teacher planning and direction. Various publisher's materials (e.g., *Language Experience in Reading* by R. V. Allen and R. Stauffer; *Sounds of Language* by W. Marting; *Math* by the Nuffield Foundation), as well as materials prepared by the teachers and the children, are available in the classroom. Field trips and walks extend the pupils' experiences. Teachers work with school psychologists to define and analyze educational problems and plan carefully defined individual solutions consistent with the TEEM approach.

The major goals for children are attended to by the teachers through a process called "orchestration." In this process, the child learns language, intellectual skills, attitudes, and societal arts and skills in a single activity. The teachers are trained to use imitation and modeling techniques as a means for developing all goal areas.

TEEM has specific goals regarding parents, including encouraging their frequent contact with school and inviting them to observe and participate in the classroom. Recently, more specified methods and approaches for implementing these goals have been developed by the sponsor.

TEEM as Realized in Follow Through

TEEM is evaluated in twelve sites: Chicasha, OK; Des Moines, IA; Lakewood, NJ; Newark, NJ; Lincoln, NB; Wichita, KS; Baltimore, MD; Vermilion Parish, LA; Durham, NC; Fort Worth, TX; Walker County, GA; and Pike County, KY.

Bank Street Model

The Model

The Bank Street Model has the immediate goal of stimulating children's cognitive and affective development, and the long range goal of effecting community change. It emphasizes personal growth of children, parents and teachers. Academic skills and emotional social development are viewed as complementary processes; both are emphasized equally. The classroom is designed to provide a stable, organized environment. Within it, children participate actively, supported by adults who help to expand their world and sensitize them to the meaning of their experiences within it. Academic skills are acquired within the broad context of direct experiences planned to provide appropriate ways of organizing and extending the children's expressed interests. Math, reading, and language are taught as tools to carry out an investigation of these interests. Children plan their learning tasks with teachers and

make autonomous choices when appropriate. A wide variety of Bank Street and commercial materials are available, such as the Bank Street readers and language stimulation materials. Children write creative stories, write their own books, read for pleasure, engage in dramatic play, music, and art. Social studies are also emphasized in the Bank Street approach.

The teachers play a vital role in this model, using diagnostic tools to analyze child behavior, child-adult interaction, and the social and physical milieu of the classroom. The staff development program aims at developing a repertoire of teaching strategies from which to choose and insights into how to enhance children's capacity to probe, reason, solve problems, and express their feelings freely and constructively. Since the teaching is based on study of the child's strengths and learning style, there is strong emphasis on individual follow-up.

The Bank Street Model as Realized in Follow Through

The Bank Street Model is evaluated in eight sites: New York, NY; Philadelphia, PA; Brattleboro, VT; Fall River, MA; New Haven, CT; Rochester, NY; Wilmington, DE; and Macon County, AL.

Direct Instruction Model University of Oregon College of Education

The Model

The Direct Instruction Model is a behaviorally oriented educational program. It utilizes a tightly controlled instructional methodology and highly structured teaching materials. Its aim is to accelerate the learning of disadvantaged children in reading, language, and arithmetic. Although the instruction is programmed, the emphasis is placed on the children's learning intelligent behavior rather than specific pieces of information by rote memorization. The Direct Instruction approach uses a fast moving series of programmed questions and answers. Teachers present specified questions to elicit a verbal child response. Proper responses are reinforced and wrong answer corrected according to specified procedures. These questions, answers, and correction procedures are contained in the Direct Instructional System in Arithmetic and Reading (Distar) materials published by Science Research Associates (SRA). Noncore subjects are generally introduced after mastery of basic skills.

Direct Instruction teachers are trained in the use of Distar programs. Teachers use these programmed materials with small homogenous groups of children for set periods of time. The groups rotate by

schedule. The children follow this group instruction with self-directed practice in workbooks. Planned home practice or new skills are also coordinated with the classroom lesson. The Direct Instruction goal for teachers is that they become proficient practitioners of the model's techniques. Criterion-referenced tests are administered to children at frequent and regular intervals to provide information to the teachers on student progress. Supervisors use video taping and observations to allow teachers to evaluate their own performances in the classroom.

Parents participate in the program in several capacities: some are employed in each classroom on a permanent basis as teacher aides (one or two per classroom) and assistants; others are employed as needed and trained to administer the criterion-referenced pupil progress tests and operate the video tape equipment to film the teacher at work in the classroom; still others are employed as family workers. In this latter capacity they acquaint parents with the Direct Instruction program, provide specially developed materials which parents can use at home to supplement classroom instruction, make available to those parents who so desire, a sponsor-developed programmed course in child management, encourage participation in PAC meetings, and assist in training the classroom aides and assistants. Finally, parent workers provide parents not directly involved in the school program with information about their child's progress and organize parents experiencing difficulties into problem-solving groups.

The Direct Instruction Model as Realized in Follow Through

The Direct Instruction Model is evaluated in ten sites: New York, NY; Grand Rapids, MI; West Iron County, MI; Flint, MI; Providence, RI; East St. Louis, IL; Racine, WI; Dayton, OH; Tupelo, MS; and Williamsburg, SC.

Behavior Analysis Model Support and Development Center for Follow Through, University of Kansas The Model

The Behavior Analysis Model (BA) recommends a highly structured yet flexible approach. Its primary objective is the children's mastery of reading, arithmetic, handwriting, and spelling skills. The program includes aspects of team teaching, non-graded classrooms, programmed instruction, individualized teaching, and a token reinforcement system. The result is an education system which unites professional educators, para-professionals, and parents in the teaching process.

As an instructional system, BA follows a standard but flexible pattern. The BA program gives primary emphasis to the basic academic skills of reading, arithmetic and language arts in the primary grades. This emphasis does not imply that music, science, art, and social studies are unimportant. It only asserts the primary importance of the core subjects as a necessary foundation for success and achievement throughout school.

The BA model is operationalized by establishing a "token economy" or "contracting arrangement" within each classroom. Teachers award tokens for improved social and academic performance. The children can use these tokens during an exchange period to purchase activities of their choosing, such as games, toys, and books. Tokens and praise are distributed according to individual rather than group performance. The sponsor deems this instructional approach appropriate for all children regardless of their socioeconomic and/or educational status.

Teachers may choose among sponsor-developed and commercial learning materials, but are encouraged to select those which can be adapted to the model. Using a machine-readable data form, teachers prepare continuous progress reports on each child. The data is then computer-analyzed and an individual progress prescription is returned within a day. Teachers are trained in the use of systematic, positive reinforcement. The sponsor supports the elimination of punitive and coercive teacher behavior and encourages teachers to set specific academic objectives for the child.

To provide the necessary amount of individual attention, BA classrooms are staffed by three or four adults. The lead teacher heads the team and generally takes special responsibility for the reading instruction. A full-time aide usually takes special responsibility for the small math groups, and the parent aide(s) concentrates on handwriting and spelling lessons and individual tutoring. At the end of eight weeks the teaching parents may continue or not as they choose. Although many parents serve only for an eight week session and teach in only one curriculum area, some teach a full year in as many as three curriculum areas. Many eventually become permanent teacher aides.

The BA Model as Realized in Follow Through

The BA Model is evaluated in eight sites: New York, NY; Philadelphia, PA; Portageville, MO; Trenton, NJ; Kansas City, MO; Louisville, KY; Waukegan, IL; and Meridan, IL.

Cognitively Oriented Curriculum Model High/Scope Educational Research Foundation

The Model

The Cognitive Curriculum Model is a developmental model, based in part on development theory and cognitive structure as defined by Piaget. The focus is on developing children's ability to reason. Goals for the individual children include development of skills in initiating and sustaining independent activity, defining and solving problems, articulating thoughts through language, assuming responsibility for decisions and actions, and working cooperatively with others to make decisions. The approach is designed to provide experiences through which children can develop their conceptual and reasoning processes, as well as their competencies in academic areas. The model provides a framework for structuring the classroom and for arranging and sequencing equipment and material in learning centers. These centers focus on math, science, reading, social studies, art and on interests such as house-keeping, construction, or puzzles. Dion reading, Nuffield and Cusinaire math, American Association for the Advancement of Science (AAAS) and Science Curriculum Improvement Study (SCIS) and science material are used. Children choose their activities and work with teachers in small groups.

Staff development is an essential component of this sponsor's model. Teachers are taught to be catalysts and motivators of children's learning, rather than skill trainers or information providers. This objective is pursued through intensive training courses which occur three times a year. These courses are designed to sensitize the teachers to the way children think and behave at different stages of development, and to supplement the sponsor-provided teacher's manual. Training and developing logical thinking skills in four major cognitive area (classification, seriation, spatial relations, and temporal relations) are a part of the teacher training program.

Central to this sponsor's model is the focus on parent involvement as an essential component of the success of the child's education. The goals for parent participation include the development of a sense of community between the school's and parents' objectives for children, and building the support for improving the fundamental parent-child relationship. Although the home visiting programs vary from community to community, the sponsor's

intent is that either the teacher or a home visitor, knowledgeable in the curriculum, will bring the essential features of this curriculum into the home. In this way, the child's learning at home can be reinforced by the parents through use of materials found in the home. Various other kinds of parent activities are also found in these communities. These activities include neighborhood meetings centered on topics such as home management, nutrition, selling and employment; providing an information network to inform parents about jobs; establishing a parent store where foods and other homemade goods can be sold or exchanged; and holding PAC meetings and various other committee meetings. This sponsor encourages parent activities that are responsive to the needs and interests of the community.

The Cognitive Model as Realized in Follow Through

The Cognitive Curriculum Model is evaluated in six sites: New York, NY; Okaloosa County, FL; Greeley, CO; Seattle, WA; Chicago, IL; and Leflore County, MS.

Florida Parent Education Model University of Florida

The Model

The Parent Education Model focuses on motivating parents to be primary educators of their children. For each class, two parents serve as teaching aides in the classroom and also visit the parents of all the children in the class, teaching them to teach their children. These parents also assist other parents with personal needs and problems.

Basic to this model is the belief that parents, since they are uniquely qualified to guide their children's emotional and intellectual development, play a critical role in their children's education. Accordingly, this sponsor seeks to motivate parents to participate directly in their child's education both in the classroom and at home. The Parent Education Model does not enunciate specific achievement goals for children, nor does it recommend a particular classroom curriculum or teaching strategy; this model focuses exclusively on involving parents as equal partners in the educational process.

The Parent Education Model uses a Parent Educator, a specially trained home worker who teaches parents to teach their children at home. (Parent Educators are themselves Follow Through Parents.) Two Parent Educators are assigned to each classroom and spend half their time as instructional teaching assistants in the classroom and half in visiting parents. Every child's home is visited bi-

monthly by a Parent Educator. During this home visit, the Parent Educator teaches the parent to work with the child in completing specially developed, individually assigned learning tasks before the next visit. These learning tasks are crucial to this model and are developed by teachers and Parent Educators with appropriate assistance from the sponsor. Learning tasks are assigned by the teacher to meet the individual child's learning needs and enrich classroom instruction. A conscientious effort is made to construct tasks using materials commonly found in the home or easily obtainable. (When necessary, materials are provided by the Parent Educator.) Tasks are often Piagetian in nature. The Parent Educator ensures that the parent thoroughly understands the task and how to use it with the child before leaving the home. During the next visit, the Parent Educator ascertains the child's response to the task and discusses the most appropriate "next step." Thus the parent becomes involved as a guiding force in the child's education. Following the home visit, the Parent Educator provides feedback to the teacher; and the two then jointly plan for the next home visit.

This partnership between home and school is reinforced by the assistance the Parent Educator provides with personal parent needs and problems. The Parent Educator is trained to make referrals for parents regarding medical, psychological, and social services and employment matters. The Parent Educator also encourages parents to join PAC and participate in other school and community activities (including classroom volunteering).

The Parent Education Model as Realized in Follow Through

The Parent Education Model is evaluated in these nine sites: Philadelphia, PA; Richmond, VA; Yakima, WA; Houston, TX; Lawrenceburg, IN; Jacksonville, FL; Jonesboro, AR; Chattanooga, TN; and Hillsborough, FL.

EDC Open Education Follow Through Program Education Development Center

The Model

The EDC Open Education approach seeks to stimulate learning by providing children with a great variety of materials and experience within a supportive emotional environment. The sponsor believes children learn at individual rates and in individual ways, and teachers should adapt approaches to encourage individual progress and responsibility in learning.

The EDC Model is predicated on the notion that learning, particularly cognitive learning, occurs best when children are offered a wide range of materials and problems to investigate within an open, supportive environment. According to this sponsor, a child's ability to learn depends in part on the opportunities and experiences provided by the educational setting. The sponsor believes that the EDC approach, derived from practices of British infant and primary schools and Piagetian research, is appropriate for all children, regardless of their socioeconomic or educational status. The EDC approach is operationalized by sponsor advisory teams who work with parents, teachers, and school administrators in each site to help realize the EDC open-education philosophy. This advisory team assists in setting up classrooms and selecting a variety of books and materials from which local educators can choose.

The sponsor believes that there is no uniform way to teach reading, writing, or arithmetic skills, and no uniform timetable for all children to follow. Children are not compared with other children and do not receive standardized tests. Consequently, EDC classrooms and teachers vary greatly. Teachers often divide classrooms into interest areas where children may work part or all of the day. Traditional subjects important in the open classroom may be combined with these interest groups. The teacher may work with the entire class, small groups, or individuals. Parents sometimes serve as classroom aides and assist in curriculum planning. In sum, the EDC Model is more a philosophy than a technique.

Since the sponsor does not prescribe a detailed instructional program and feels that the open classroom philosophy is appropriate for all voluntary teachers, this model demands a highly creative and resourceful teacher and is perhaps the most teacher-dependent of the Follow Through models. Teachers must diagnose each child's strengths, potential, and interests and then strive to provide instructional units reflecting that information. They are trained to provide a "hidden structure," to act as guides and resources, to make suggestions and to give encouragement, as the primary methods of extending their pupils' learning activities. Within this environment the pupils are encouraged to work at their own pace, learn from one another, and make choices about their own work.

The parents in this model are encouraged to become involved. Their primary involvement is through their work on the advisory teams and in the

PAC organizations. This model's goal for parents is to help them "grow" and understand the concepts of open education. Its general approach is cognitive, with an almost equally heavy socioemotional emphasis. Although there is some stress on specific academic skills, the foci of this model are learning how to learn, developing an appreciation for learning, and encouraging children to take responsibility for their own learning.

The EDC Model as Realized in Follow Through

The EDC Model is evaluated in the eight sites: Philadelphia, PA; Burlington, VT; Lackawanna County, PA; Morgan Community School in Washington, DC; Patterson, NJ; Chicago, IL; Laurel, DE; and Johnston County, NC.

Volume IV-A:

An Evaluation of Follow Through

The Follow Through models place varying degrees of emphasis on the acquisition of basic skills, cognitive conceptual skills, and affective development. Although all sponsors expected to demonstrate effectiveness in all domains by the end of third grade, we can expect the models to produce various time sequences of progress in achieving this goal. We have divided the progress of these children during the course of the program into two parts: progress during kindergarten and first grade (early) and progress during second and third grade (late). A study of the progress of FT children during these two intervals shows that most programs produce substantial progress early on math measures. However, only a few of the programs are able to maintain these early benefits in math during the later period of the program.

The reading area appears to be much less tractable. Direct Instruction, Behavior Analysis, and Bank Street models produce predominately non-negative effects, that is, progress in reading which is either greater than or equal to the progress of comparison children. Only the children associated with the Direct Instruction Model appear to perform above the expectation determined by the progress of the non-Follow Through children. Moreover, the Direct Instruction children are the only group which appears to make more progress in reading, both early and late. In general, most models appear to be more effective during kindergarten and first grade than during second and third grade. The Direct Instruction Model is the only program which consistently produces substantial progress.

Abt Associates' Final Follow Through Reports

•Volume IV-A: An Evaluation of Follow Through (Office of Education Series Vol. II-A)
Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. Abt Associates Report No. AAI-76-196A under USOE Contract No. 300-75-0134. April 15, 1977.

Contains a description of the study, the educational approaches examined, a discussion of the analytic strategies and methods of presenting results, along with a summary of the results.

•Volume IV-B: Effects of Follow Through Models (Office of Education Vol. II-B)
Bock, G., & Stebbins, L. B., with E. C. Proper. Abt Associates Report No. AAI-76-196B under USOE Contract No. 300-75-0134. April 15, 1977.

Contains a comprehensive discussion of the results.



Phyllis Anderson Wilken

*Former elementary principal
and educational consultant
Center for the Study of Reading
College of Education
University of Illinois*

Foreword by P. David Pearson

Turning Our School Around

Seven Commonsense Steps to School Improvement

"Phyllis Wilken describes how she worked as the principal with the teachers, parents and students to transform a troubled school into an award winning one. She mixes the best of the old with the best of the new by calling upon her experience as a beginning teacher in a one-room rural school and combines them with knowledge of educational organization, staff development, and curriculum planning. The result is a clearly written and uncommonly sensible handbook for anyone seriously interested in school change.

Jean Osborn, Associate
Director of Center for the
Study of Reading,
University of Illinois

Turning Our School Around: Seven Commonsense Steps to School Improvement
is now available through ADI.

See page 70 in this issue for ordering information.

A Constructive Look at Follow Through Results

Carl Bereiter, Ontario Institute for Studies in Education
Midian Kurland, University of Illinois at Urbana-Champaign

Reprinted from Interchange, Vol.12, Winter, 1981, with permission

Follow Through is a large compensatory education program that operated in scores of communities across the United States throughout the seventies and that continues, on a reduced scale, today. During its most active phase, it was conducted as a massive experiment, involving planned variation of education approaches and collection of uniform data at each site. The main evaluation of outcomes was carried out by Abt Associates, Inc. (a private consulting firm, based in Cambridge, Massachusetts) on the second and third cohorts of children who reached third grade in the program, having entered in kindergarten or first grade. In a series of voluminous reports, Abt Associates presented analyses indicating that among the various education approaches tried, only those emphasizing "basic skills" showed positive effects when compared to Non-Follow Through treatments. House, Glass, McLean, and Walker (1978a) published a critique of the Abt Associates evaluation, along with a small reanalysis that found essentially no significant differences in effectiveness among the planned variations in educational approaches. Because of the great social importance attached to educational programs for disadvantaged groups and because no other large-scale research on the topic is likely to materialize in the near future, the Follow Through experiment deserves continuing study. The study reported here is an attempt, through more sharply focused data analysis, to obtain a more definitive answer to the question of whether different educational approaches led to different achievement outcomes.

Is it possible that the Follow Through planned variation experiment has yielded no findings of value? Is it possible, after years of effort and millions of dollars spent on testing different approaches, that we know nothing more than we did before about ways to educate disadvantaged children? This is the implicit conclusion of the widely publicized critique by House, Glass, McLean, and Walker (1978a, 1978b). House et al found no evidence that the various Follow Through models differed in effectiveness from one another or from Non-Follow Through programs. The only empirical finding House et al were willing to credit was that there was great variation in results from one Follow Through site to another. This conclusion, as we shall show, is no more supportable than the conclusions House et al rejected. Accordingly, if we were to follow House, Glass, McLean, and Walker's lead, we should have to conclude that there are no substantive findings to be gleaned from the largest educational experiment ever conducted.

It would be a serious mistake, however, to take the critique by House et al as any kind of authoritative statement about what is to be learned from Follow Through. The committee assembled by

House was charged with reviewing the Abt Associates evaluation of Follow Through (Stebbins et al, 1977), not with carrying out an inquiry of their own. More or less, the committee stayed within the limits of this charge, criticizing a variety of aspects of the design, execution, and data analysis of the experiment. Nowhere in their report do the committee take up the constructive problem that Abt Associates had to face or that any serious inquiries will have to face. Given the weaknesses of the Follow Through experiment, how can one go about trying to extract worthwhile findings from it?

In this paper we try to deal constructively with one aspect of the Follow Through experiment: the comparison of achievement test results among the various sponsored approaches. We try to show that if this comparison is undertaken with due cognizance of the limitations of the Follow Through experiment, it is possible to derive some strong, warranted, and informative conclusions. We do not present our research as a definitive, and certainly not as a complete, inquiry into Follow Through results. We do hope to show, however, that the conclusion implied by the House committee—that the Follow Through experiment is too flawed to yield any positive findings—is gravely mistaken.

Delimiting the Problem

Although Project Follow Through has numerous shortcomings as an experiment, the seriousness of these shortcomings varies greatly depending on what questions are asked of the data. One shortcoming was in the outcome measures used, particularly in their limited range compared to the range of objectives pursued by Follow Through sponsors. The House committee devotes the largest part of its critique to this shortcoming, although it is a shortcoming that limits only the range of conclusions that may be drawn. House et al allow, for instance, that the Metropolitan Achievement Test was "certainly a reasonable choice for the material it covers" (1978a, p. 138). Accordingly, Follow Through's shortcomings as to outcome measures ought not to stand in the way of answering questions that are put in terms appropriate to the measures that were used.

Is it possible that the Follow Through planned variation experiment has yielded no findings of value? Is it possible, after years of effort and millions of dollars spent on testing different approaches, that we know nothing more than we did before about ways to educate disadvantaged children?

Another shortcoming, recognized by all commentators on Follow Through, is the lack of strictly comparable control groups. Follow Through and Non-Follow Through groups at the same site differed from one another in uncontrolled and only partly measurable ways, and the differences themselves varied from site to site. This circumstance makes it difficult to handle questions having to do with whether children benefited from being in Follow Through, because such questions require using Non-Follow Through data as a basis for inferring how Follow Through children would have turned out had they not been in Follow Through.

Much of the bewildering complexity of the Abt Associates' analyses results from attempts to make up statistically for the lack of experimental comparability. We do not intend to examine those attempts except to note one curiosity. The difficulty of evaluating "benefits" holds whether one is asking about the effects of Follow Through as a whole, the effects of a particular model, or the effect of a Follow

Through program at a single site. The smaller the unit, however, the more vulnerable the results are likely to be to a mismatch between Follow Through and Non-Follow Through groups. On the one hand, to the extent that mismatches are random, they should tend to average out in larger aggregates. On the other hand, at a particular site, the apparent success or failure of a Follow Through program could depend entirely on a fortuitously favorable or unfavorable match with a Non-Follow Through group.

For unknown reasons, both the Abt Associates and the House committee analysts have assumed the contrary of the point just made. While acknowledging, for instance, that the prevalence of achievement test differences in favor of Non-Follow Through groups could reflect mismatch, they are able to make with confidence statements like "Seven of the ten Direct Instruction sites did better than the comparison classes but three of the Direct Instruction sites did worse" (House et al, 1978a, p. 154). Such a statement is nonsense unless one believes that at each of the ten sites a valid comparison between Follow Through and Non-Follow Through groups could be made. But if House et al believe that, how could they then believe that the average of those ten comparisons is invalid? This is like arguing that IQ tests give an invalid estimate of the mean intelligence level of disadvantaged children and then turning around and using those very tests to classify individual disadvantaged children as retarded.

There is an important class of questions that may be investigated, however, without having to confront the problem of comparability between Follow Through and Non-Follow Through groups. These are questions involving the comparison of Follow Through models with one another. A representative question of this kind would be—how did the Follow Through models compare with one another in reading achievement test scores at the end of third grade? There are problems in answering such a question, but the lack of appropriate control groups is not one of them. We can, if we choose, simply ignore the Non-Follow Through groups in dealing with questions of this sort.

Questions about the relative performance of different Follow Through models are far from trivial. The only positive conclusions drawn by Abt Associates relate to questions of this kind, and the House committee's report is largely devoted to disputing those conclusions—that is, disputing Abt's conclusions that Follow Through models emphasizing basic skills achieved better results than others in basic skills and in self-concept. The models represented in Follow Through cover a wide range of educa-

Although Project Follow Through has numerous shortcomings as an experiment, the seriousness of these shortcomings varies greatly depending on what questions are asked of the data.

tional philosophies and approaches to education. Choose any dimension along which educational theories differ and one is likely to find Follow Through models in the neighborhood of each extreme. This is not to say that the Follow Through models are so well distinguished that they provide clean tests of theoretical issues in education. But the differences that are there—like, for instance, the difference between an approach based on behavior modification principles and an approach modeled on the English infant school—offer at least the possibility of finding evidence relevant to major ideological disputes within education.

Unscrambling the Methodology

The Abt Associates analysts were under obligation to try to answer the whole range of questions that could be asked about Follow Through effects. In order to do this in a coherent way, they used one kind of statistic that could be put to a variety of uses. This is the measure they called "effect size," an adjusted mean difference between the Follow Through and Non-Follow Through subjects at a site. Without getting into the details of how effect size was computed, we may observe that this measure is more suitable for some purposes than for others. For answering question about benefits attributable to Follow Through, some such measure as effect size is necessary. For comparing one Follow Through model with another, however, the effect size statistic has the significant disadvantage that unremoved error due to mismatch between a Follow Through and Non-Follow Through group is welded into the measure itself. As we noted in the preceding section, comparisons of the effectiveness of Follow Through models with one another do not need to involve Non-Follow Through data. Because effect size measures will necessarily include some error due to mismatch (assuming that covariance adjustments cannot possibly remove all such error), these measures will contain "noise" that can be avoided when making comparisons among Follow Through models.

The Abt Associates analysts used several different ways of computing effect size, the simplest of

which is called the "local" analysis. This method amounts to using the results for each cohort of subjects at each site as a separate experiment, carrying out a covariance analysis of Follow Through and Non-Follow Through differences as if no other sites or cohorts existed. Although this analysis has a certain elegance, it clearly does not take full advantage of the information available; the "pooled" analysis used by Abt, which uses data on whole cohorts to calculate regression coefficients and at the same time includes dummy variables to take care of site-specific effects, is much superior in this respect. The House committee, however, chose to use effect size measures based on the "local" analysis in their own comparison of models. In doing so, they used the least powerful of the Abt effect size measures, all of which are weakened (to unknown degrees) by error due to mismatch.

In their comparisons of Follow Through models, Abt Associates analysts calculated the significance of effects at different sites, using individual subjects at the sites as the units of analysis, and then used the distribution of significant positive and negative effects as an indicator of the effectiveness of the models. The House committee argued, on good grounds we believe, that the appropriate unit of analysis should have been sites rather than individual children. To take only the most obvious argument on this issue, the manner of implementing a Follow Through model is a variable of great presumptive significance, and it is most reasonably viewed as varying from site to site rather than from child to child. Having made this wise decision, however, the House committee embarked on what must be judged either an ill-considered or an excessively casual reanalysis of Follow Through data.

Although the reanalysis of data by the House committee occupies only a small part of their report and is presented by them with some modesty, we believe their reanalysis warrants severe critical scrutiny. Without that reanalysis, the House committee's report would have amounted to nothing more than a call for caution in interpreting the findings of the Abt Associates analysts. With the reanalysis, the House committee seems to be declaring that there are no acceptable findings to be interpreted. Thus a great deal hinges on the credibility of their reanalysis.

Let us therefore consider carefully what the House committee did in their reanalysis. First, they used site means rather than individual scores as the unit of analysis. This decision automatically reduced the Follow Through planned variation experiment from a very large one, with an N of thousands, to a rather small one, with an N in the neighborhood of one

hundred. As previously indicated, we endorse this decision. However, it seems to us that when one has opted to convert a large experiment into a small one, it is important to make certain adjustments in strategy. This the House committee failed to do. If an experiment is very large, one can afford to be cavalier about problems of power, since the large N will presumably make it possible to detect true effects against considerable background noise. In a small experiment, one must be watchful and try to control as much random error as possible in order to avoid masking a true effect.

However, instead of trying to perform the most powerful analysis possible in the circumstances, the House committee weakened their analysis in a number of ways that seem to have no warrant. First, they chose to compare Follow Through models on the basis of Follow Through/Non-Follow Through differences, thus unnecessarily adding error variance associated with the Non-Follow Through groups. Next, they chose to use adjusted differences based on the "local" analysis, thus maximizing error due to mismatch. Next, they based their analysis on only a part of the available data. They excluded data from the second kindergarten-entering cohort, one of the largest cohorts, even though these data formed part of the basis for the conclusions they were criticizing. This puzzling exclusion reduced the number of sites considered, thus reducing the likelihood of finding significant differences. Finally, they divided each effect-size score by the standard deviation of test scores in the particular cohort in which the effect was observed. This manipulation served no apparent purpose. And minor though its effects may be, such as they are would be in the direction of adding further error variance to the analysis.

The upshot of all these methodological choices was that, while the House group's reanalysis largely confirmed the ranking of models arrived at by Abt Associates, it showed the differences to be small and insignificant. Given the House committee's methodology, this result is not surprising. The procedures they adopted were not biased in the sense of favoring one Follow Through model over another; hence it was to be expected that their analysis, using the same effect measures as Abt, would replicate the

Instead of trying to perform the most powerful analysis possible in the circumstances, the House committee weakened their analysis in a number of ways that seem to have no warrant.

rankings obtained by Abt. (The rank differences shown in Table 7 of the House report are probably mostly the result of the House committee's exclusion of data from one of the cohorts on which the Abt rankings were based.) On the other hand, the procedures adopted by the House committee all tended in the direction of maximizing random error, thus tending to make differences appear small and insignificant.

The analysis to be reported here is of the same general type as that carried out by the House committee. Like the House committee, we use site means rather than scores for individuals as the unit of analysis. The differences in procedure all arise from our effort to minimize random error and thus achieve the most powerful analysis possible. The following are the main differences between our analysis and the House et al analysis:

1. We used site means for Follow Through groups as the dependent variable, using other site-level scores as covariates. The House committee used locally adjusted site-level differences between Follow Through and Non-Follow Through groups as the dependent variable, with covariance adjustments having been made on an individual basis. Our procedure appears to have been endorsed in advance by the House committee. They state: "For the sake of both inferential validity and proper covariance adjustment, the classroom is the appropriate unit of analysis" (House et al, 1978a, p. 153). While the House committee followed their own prescription in using site-level scores as dependent variables, they failed to follow it when it came to covariance adjustments.
2. When we used Non-Follow Through scores, we entered them as covariates along with other covariates. The procedure adopted by the House committee amounted, in effect, to arbitrarily assigning Non-Follow Through mean scores a regression weight of 1 while giving all other variables empirically determined regression weights. We could not see any rational basis for such a deviation from ordinary procedures for statistical adjustment.
3. We combined all data from one site as a single observation, regardless of cohort. The House committee appear to have treated different cohorts from the same site

as if they were different sites. This seemed to us to violate the rationale for analyzing data at the site level in the first place.

4. We restricted the analysis to models having data on 6 or more sites. To include in the analysis models having as few as 2 sites, as the House committee did, would, it seemed to us, reduce the power of the statistical tests to an absurd level.

The data analysis that followed from the above-mentioned decisions was quite straightforward and conventional. The dependent variable was always the mean score for a site on one or more Metropolitan Achievement Test subtests, averaged over all subjects in cohorts II and III for whom data were reported in the Abt Associates reports. Models, which ranged from 12 to 6 in number of sites, were compared by analysis of covariance, using some or all of the following covariates:

- SES—An index of socio-economic status calculated by Abt for each cohort at each site. When more than one cohort represented a site, an n-weighted mean was computed.
- EL—An index of ethnic and linguistic difference from the mainstream—treated in a manner similar to SES.

WRAT—Wide-range Achievement Test, administered near time of entry to Follow Through students. Taken as a general measure of academic readiness.

NFT—Mean score of local Non-Follow Through students on the dependent variable under analysis. As a covariate, NFT scores may be expected to control for unmeasured local or regional characteristics affecting scholastic achievement.

Two other covariates were tried to a limited extent: Raven *Progressive Matrices* scores (which, though obtained after rather than before treatment, might be regarded as primarily reflecting individual differences variance not affected by treatment) and a score indicating the number of years of Follow Through treatment experienced by subjects at a site (most Follow Through groups entered in kindergarten, thus receiving four years of Follow Through treatment; but some entered in first grade and received only three years). Our overall strategy for use of analysis of covariance was as follows: recognizing that reasonable cases could be made for and against the use of this covariate or that, we would try various combinations and, in the end, would take seriously only those results that held up over a variety of reasonable covariate sets.

Table 1
Significant Differences Between Adjusted Means (Newman-Keuls Tests, $\alpha = .05$) — "Full"
Covariance Analysis
Lower Scoring Model

| | Higher Scoring Model | | | | | | | |
|--------|-------------------------------|----------------------------|-------------------|---------------|-------------------|--------------|--------------|---------------------------------|
| | Resp. Educ. | TEEM (Tucson) | Bank Street | Direct Instr. | Behavior Analysis | Cog. Curric. | Parent Educ. | EDC Open Educ. |
| R.E. | | | | | | | | |
| TEEM | | | | | | | | |
| B.S. | | | | | | | | |
| D.I. | M1 TL M2 LA M3 TM LB | M1 TL M2 LA TM LB | M1 TL TM LB | | TL | M1 TL TM | M1 TL TM | M1 TL R M2 M3 TM LA LB |
| B.A. | S M2 R TM | S WK R TM | TR | | | | | S R M2 TR WK TM |
| C.C. | | | | | | | | |
| P.E. | | | | | | | | |
| E.D.C. | | | | | | | | |

Note: Initials in boxes indicate tests on which significant differences were obtained. The higher-scoring model is indicated by the row, lower-scoring model by the column in which the box is found. Key to tests:

- SP — Spelling
WK — Word Knowledge
R — Reading
LA — Language (Part A)
LB — Language (Part B)
M1 — Math. Computation
M2 — Math. Concepts
M3 — Math. Problem Solving
TL — Total Language
TR — Total Reading
TM — Total Math.

Table 2

| | Lower Scoring Model | | | | | | | |
|--------|-------------------------------|-------------------------------|-------------------------------|---------------|-------------------------------|-------------------------------|-------------------------------|---|
| | Resp. Educ. | TEEM (Tucson) | Bank Street | Direct Instr. | Behavior Analysis | Cog. Curric. | Parent Educ. | EDC Open Educ. |
| R.E. | | | | | | | | |
| TEEM | | | | | | | | |
| B.S. | | | | | | | | |
| D.I. | M1 LA M2 TL M3 TM LB | M1 LA M2 TL M3 TM LB | M1 LA M2 TL M3 TM LB | | M1 LA M2 TL M3 TM LB | M1 LA M2 TL M3 TM LB | M1 LA M2 TL M3 TM LB | S M1 TR WK M2 TL R M3 TM LA LB |
| B.A. | S M1 WK | S | S M1 | | | M1 | S M1 | S M1 TR WK M2 TL R M3 TM |
| C.C. | | | | | | | | |
| P.E. | | | | | | | | |
| E.D.C. | | | | | | | | |

Note: Initials in boxes indicate tests on which significant differences were obtained. The higher-scoring model is indicated by the row, lower-scoring model by the column in which the box is found.
Key to tests:

SP — Spelling
WK — Word Knowledge
R — Reading
LA — Language (Part A)
LB — Language (Part B)
M1 — Math. Computation
M2 — Math Concepts
M3 — Math. Problem Solving
TL — Total Language
TR — Total Reading
TM — Total Math.

Results

Differences in achievement test performance—Two analyses of covariance will be reported here, with others briefly summarized. Figure 1 displays adjusted and standardized means from what we call the "full" analysis of covariance—that is, an analysis using the four main covariates (SES, EL, WRAT, and NFT) described in the preceding section. The virtue of this analysis is that it controls for all the main variables that previous investigators have tried, in one way or other, to control for in comparing Follow Through models.

Table 1 notes pair-wise differences which are significant at the .05 level by Newman-Keuls tests. Figure 2 and Table 2 show comparable data for what we call the "conservative" analysis. This analysis is conservative in the sense that it eliminates covariates for which there are substantial empirical and/or rational grounds for objection. Grounds for objecting to the NFT variable as a covariate have been amply documented in Abt reports and echoed in the report of the House committee (House et al, 1978a); they will not be repeated here. Use of WRAT as a covariate has been objected to on grounds that it is not, as logically required, antecedent to treatment (Becker & Carnine, Ref. Note 1)—that is, the WRAT, though nominally a pretest, was in fact administered at a time when at least one of the models had already purportedly taught a significant amount of the content touched on by the WRAT. While we

would not suppose the SES and EL variables to be above reproach, we have not encountered criticisms suggesting their use would seriously bias results—whereas not to control for these variables would unquestionably leave the results biased in favor of models serving less disadvantaged populations. Accordingly, we have chosen them as the conservative set of covariates.

Other analyses, not reported, used different combinations of covariates from among those mentioned in the preceding section. In every case, these analyses yielded adjusted scores intermediate between those obtained from the "full" and the "conservative" analyses. Consequently, the results shown in Figures 1 and 2 may be taken to cover the full range of those observed.

In every analysis, differences between models were significant at or beyond the .05 level on every achievement variable—almost all beyond the .01 level. As Figures 1 and 2 show, models tended to perform about the same on every achievement variable. Thus there is little basis for suggesting that one model is better at one thing, another at another.

The relative standing of certain models, particularly the Tucson Early Education Model, fluctuated considerably depending on the choice of covariates.¹ Two models, however, were at or near the top on every achievement variable, regardless of the covariates used; these were Direct Instruction and Behavior Analysis. Two models were at or near the

bottom on every achievement variable, regardless of the covariates used; these were the EDC Open Education Model and Responsive Education. Differences between the two top models and the two bottom models were in most cases statistically significant by Newman-Keuls tests.

Variability between sites—The only empirical finding that the House committee was willing to credit was that there was enormous variability of effects from site to site within Follow Through models. In their words: "Particular models that worked well in one town worked poorly in another. Unique features of the local settings had more effect on achievement than did the models" (House et al, 1978, p. 156). This conclusion has recently been reiterated by the authors of the Abt evaluation report (St. Pierre, Anderson, Proper, & Stebbins, 1978) in almost the same words.

The ready acceptance of this conclusion strikes us as most puzzling. It is conceivable that all of the variability between sites within models is due to mismatch between Follow Through and Non-Follow Through groups. This is unlikely, of course, but some of the variability between sites must be due to this factor, and unless we know how much, it is risky to make statements about the real variability of effects. Furthermore there is, as far as we are aware,

no evidence whatever linking achievement to "unique features of the local setting." This seems to be pure conjecture—a plausible conjecture, no doubt, but not something that should be paraded as an empirical finding.

Our analyses provide some basis for looking at the between-site variability question empirically. Follow Through sites varied considerably in factors known to be related to achievement—socioeconomic status, ethnic composition, WRAT pretest scores, etc. To say that the variance in achievement due to these factors was greater than the variance due to model differences may be true but not very informative. It amounts to nothing more than the rediscovery of individual differences and is irrelevant to the question of how much importance should be attached to variation among Follow Through models. To say that differences in educational method are trivial because their effects are small in comparison to the effect of demographic characteristics is as absurd as saying that diet is irrelevant to children's weight because among children weight variations due to diet are small in comparison to weight variations due to age.

The variability issue may be more cogently formulated as follows: considering only the variance in achievement that *cannot* be accounted for by demo-

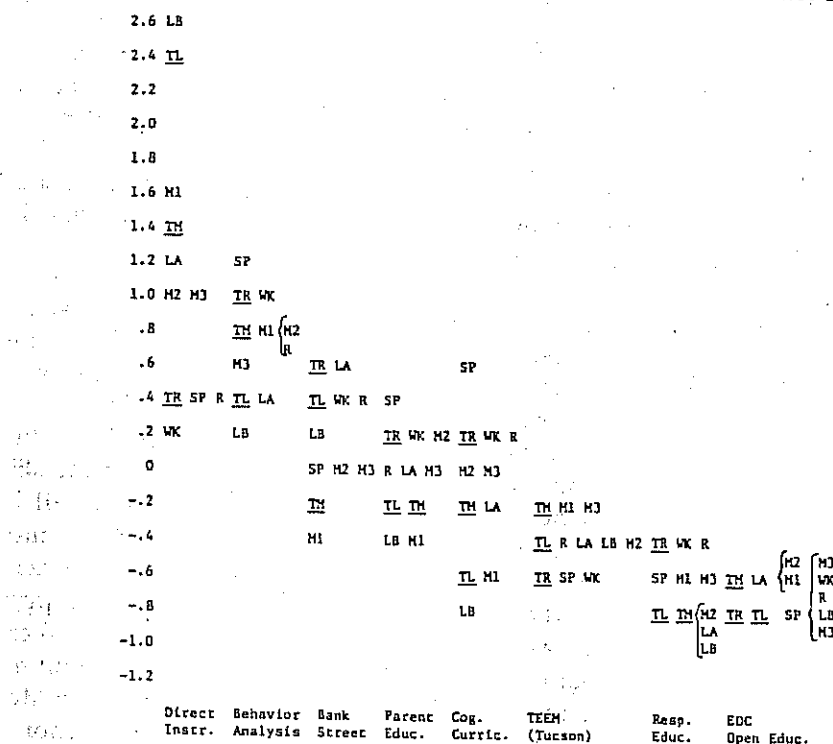


Figure 1

Standardized adjusted mean Metropolitan Achievement Test scores obtained from "full" covariance analysis (rounded to the nearest even tenth).

graphic and other entering characteristics of students, what part of that variance can be explained by differences in Follow Through models and what part remains unexplained? Our analyses provide an approximate answer to this question, since covariance adjustments act to remove variance among sites due to entering characteristics. Depending on the achievement test variable considered and on the covariates used, we found model differences to account for roughly between 17 and 55 per cent of the variance not attributable to covariates (as indexed by ω^2).

These results are shown graphically in Figures 1 and 2. Adjusted mean scores are displayed there in units of the standard deviation of residual site means. Thus, to take the most extreme case, in Figure 2 the adjusted mean score of Direct Instruction sites on Language Part B is 3.6 standard deviations above the adjusted mean score of EDC Open Education sites—that is, 3.6 standard deviations of between-site residual variability; in other words, an enormous difference compared to differences between sites within models. That is the most extreme difference, but in no case is the adjusted difference between highest and lowest model less than 1.4 standard deviations. Although what constitutes a “large” effect must remain a matter of judgment, we know of no prece-

dent according to which treatment effects of this size could be considered small in relation to the unexplained variance.

Treatment effects on other variables—Although the principal concern of this study was with achievement test differences, the method of analysis is adaptable to studying differences in other outcomes as well. Accordingly we ran several briefer analyses, looking at what Abt Associates call “cognitive / conceptual” and “affective” outcomes.

Two kinds of measures used in the Follow Through evaluation were regarded by Abt Associates as reflecting “cognitive / conceptual” outcomes—Raven’s *Progressive Matrices* (a nonverbal intelligence test) and several Metropolitan subtests judged to measure indirect cognitive consequences of learning. The House committee objected to *Progressive Matrices* on grounds that is insensitive to school instruction. This rather begs the question of effects of cognitively-oriented teaching, however. True, *Progressive Matrices* performance may be insensitive to ordinary kinds of school instruction, but does that mean it will be insensitive to novel instructional approaches claiming to be based on cognitive theories and declaring such objectives as “the ability to reason” and “logical thinking skills in four major cognitive areas (classification, seriation, spatial re-

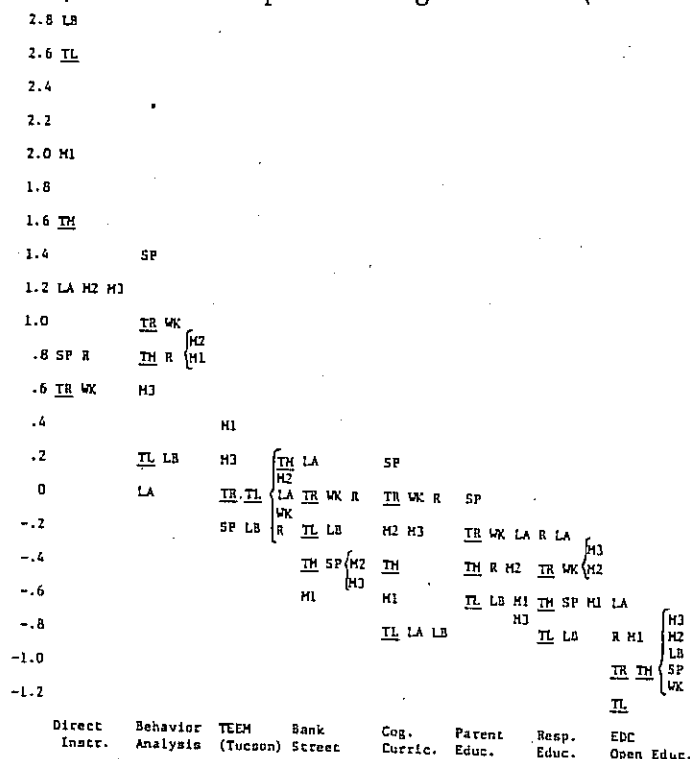


Figure 2

Standardized adjusted mean Metropolitan Achievement Test scores obtained from “conservative” covariance analysis (rounded to the nearest even tenth).

lations and temporal relations)? It seems that this should be an empirical question.

If it is an empirical question, the answer is negative. Using the same kinds of covariance analyses as were used on the achievement test variables, we found no statistically significant differences between Follow Through models in *Progressive Matrices* performance. This finding is consistent with the Abt Associates' analyses, which show few material effects on this test, and more negative than positive ones.

Among Metropolitan subtests the most obviously "cognitive" are Reading (which is, in effect, paragraph comprehension) and Mathematics Problem-Solving. As indicated in Figures 1 and 2, our analyses show differences among models on these subtests that are similar in trend to those found on the other subtests. They tend, however, to be of lesser magnitude. The most obvious explanation for the lesser magnitude of difference on these subtests is the same as that offered by House et al for the absence of differences on *Progressive Matrices*—that these subtests, reflecting more general differences in intellectual ability, are less sensitive to instruction. There is, however, a further hypothesis that should be tested. Conceivably, certain models—let us say those that avowedly emphasize "cognitive" objectives—are doing a superior job of teaching the more cognitive aspects of reading and mathematics, but the effects are being obscured by the fact that performance on the appropriate subtests depends on mechanical proficiency as well as on higher-level cognitive capabilities. If so, these hidden effects might be revealed by using performance on the more "mechanical" subtests as covariates.

This we did. Model differences in Reading (comprehension) performance were examined, including Word Knowledge as a covariate. Differences in Mathematics Problem Solving were examined, including Mathematics Computation among the covariates. In both cases the analyses of covariance revealed no significant differences among models. This is not a surprising result, given the high correlation among Metropolitan subtests. Taking out the variance due to one subtest leaves little variance in another. Yet it was not a forgone conclusion that the results would be negative. If the models that proclaimed cognitive objectives actually achieved those objectives, it would be reasonable to expect those achievements to show up in our analyses.

The same holds true for performance on the affective measures included in the Follow Through evaluation. The Abt Associates' analyses show that the ranking of models on affective measures corresponds

closely to their ranking on achievement measures. House et al point out, however, that the instruments used place heavy demands on verbal skills. Conceivably, therefore, if reading ability were controlled statistically, the results might tell a different story. We analyzed scores on the Coopersmith Self-Concept Inventory, including reading subtest scores along with the other covariates. The result showed no significant difference among models on the Coopersmith. This finding could mean either that there are no differences between models in effects on self-concept or that self-concept among disadvantaged third-graders is sufficiently dependent on reading ability that, when one statistically removes reading ability differences, one at the same time removes genuine self-concept differences. We know of no way to resolve this ambiguity with the available data. One thing is clear, however: removing effects due to reading achievement does not in any way yield results either favoring models that emphasize self-concept or disfavoring models that emphasize academic objectives.

Discussion

Before attempting to give any interpretation of Follow Through results, we must emphasize the main finding of our study—that *there were results*. Follow Through models were found to differ significantly on every subtest of the Metropolitan Achievement Test.

Let us briefly compare our findings with those of Abt Associates and the House committee.

1. We disagree with both Abt and House et al in that we do 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.
2. Our findings on the ranking of Follow Through models on achievement variables are roughly in accord with those of the House Committee, but we differ from the House committee in finding significant differences among models on all achievement variables whereas they found almost none. The similarities are no doubt due to the fact

that the two analyses used the same basic units—site-level means. The difference in significance of outcomes is apparently due to the variety of ways (previously discussed) in which our analysis was more powerful than theirs.

3. The Abt Associates' results indicate that among major Follow Through models, there is only one "winner" in the sense of having a preponderance of positive effects—namely, Direct Instruction. All other models showed predominately null or negative effects. Our results are not exactly comparable in that we compared Follow Through models only with one another and not with Non-Follow Through groups; consequently we cannot speak of "positive" or "negative" effects. However, our results show two models to be above average on all achievement subtests and two models to be below average on all subtests. Thus our results may be said to indicate two "winners"—Direct Instruction and Behavior Analysis—and two "losers"—EDC Open Education and Responsive Education.

We put the words "winners" and "losers" in quotation marks because, of course, Follow Through was not a contest with the object of attaining the highest possible achievement test scores. It simply happens that the outcomes on which Follow Through models are found to differ are achievement test scores. That other criteria might have shown different winners and losers (a point heavily emphasized by the House committee) must remain a conjecture for which all the available evidence is negative. What we have are achievement test differences, and we must now turn to the question of what those differences might mean.

It lies outside the scope of this paper to discuss the importance of scholastic achievement itself. The more immediate issue is whether the observed differences in achievement test scores reflect actual differences in mastery of reading, mathematics, spelling, and language.

One obvious limitation that must be put on the results is that the Metropolitan Achievement Test, like all other standardized achievement batteries, covers less than the full range of achievement objectives. As House et al point out, the test does not cover "even such straightforward skills as the ability to read aloud, to write a story, or to translate an ordinary problem into numbers" (1978b, p. 473).

This much is certainly true, but House et al then go on to say, "it would be reckless to suppose that the results of the testing indicate the attainment of these broader goals" (p. 473). "Reckless" is far too strong a word here.² From all we know about the intercorrelation of scholastic skills, one could be fairly confident in assuming that children who perform above average on the MAT would also perform above average on tests of the other skills mentioned. A glance again at Figures 1 and 2 tells us that achievements in a variety of areas tend to go together. Given the homogeneous drift of scores downward from left to right in those figures, it is hard to imagine another set of achievement measures in mathematical and language skills that would show a trend in the opposite direction. Such a trend cannot be declared impossible, of course, but if House et al expect us to take such a possibility seriously, then they ought to provide some evidence to make it plausible.

A more serious kind of charge is that the MAT is biased in favor of certain kinds of programs. If true, this could mean that the observed test score differences between models reflect test bias and not true differences on the achievement variables that the test is supposed to measure. We must be very careful, however, in using the term *bias*. One sometimes hears in discussions of Follow Through statements that the MAT is biased in favor of models that teach the sort of content measured by the MAT. This is a dangerous slip in usage of the word *bias* and must be avoided. It makes no sense whatever to call it bias when an achievement test awards higher scores to students who have studied the domain covered by the test than to students who have not. It would be a very strange achievement test if it did not.

It is meaningful, however, to say that an achievement test is biased in its sampling of a domain of content, but even here one must be careful not to abuse the term. The Mathematics Concept subtest of the MAT, for instance, is a hodge-podge of knowledge items drawn from "old math," "new math," and who knows what. For any given instructional program, it will likely be found that the test calls for knowledge of material not covered by that program—but that doesn't mean the test is biased against the program. The test obviously represents a compromise that cannot be fully satisfactory to any program. The only ground for a charge of bias would be that the compromise was not even-handed. Investigating such a charge would require a thorough comparison of content coverage in the test and content coverage in the various Follow Through programs. It does no good to show that for a particu-

lar program there are discrepancies between content covered and content tested. The same might be equally true of every program.

As far as the Follow Through evaluation goes, the only MAT subtest to which a charge of content bias might apply (we have no evidence that it does) is Mathematics Concepts. The other subtests all deal with basic skills in language and mathematics. Different programs might teach different methods of reading or doing arithmetic, and they might give different amounts of emphasis to these skills, but the skills tested on the MAT are all ones that are appropriate to test regardless of the curriculum. Even if a particular Follow Through model did not teach arithmetic computation at all, it would still be relevant in an assessment of that program to test students' computational abilities; other people care about computation, even if the Follow Through sponsor does not. The reason why Mathematics Concepts may be an exception is that, while everyone may care about mathematical concepts, different people care about different ones, and so a numerical score on a hodgepodge of concepts may not be informative.

While such skill tests as those making up the bulk of the MAT are relatively immune to charges of content bias, they can be biased in other ways. They may, perhaps, be biased in the level of cognitive functioning that they tap within a skill area. The House committee implies such a bias when they say, "the selection of measures favors models that emphasize rote learning of the mechanics of reading, writing, and arithmetic" (House et al, 1978a, p. 14S). This is a serious charge and, if true, would go some way toward discrediting the findings.

But House et al offer no support for this charge, and on analysis it seems unlikely that they could. Their statement rests on three assumptions for which we know of no support: (1) That "the mechanics of reading, writing, and arithmetic" can be successfully taught by rote; (2) that there were Follow Through models that emphasized rote learning (the model descriptions provided by Abt give no suggestion that this is true)³ and (3) that the MAT measures skills in such a way that the measurement favors children who have learned those skills by rote rather than through a meaningful process. We must conclude, in fact, that since the House committee could not have been so naive as to hold all three of these assumptions, they must have introduced the word "rote" for rhetorical effect only. Take the word out and their statement reduces to an unimpressive complaint about the limited coverage of educational objectives in the Follow Through evaluation.

A final way in which skill tests might be biased is in the *form* of the test problems. Arithmetic compu-

tation problems, for instance, might be presented in notation that was commonly employed in some programs and not in others; or reading test items might use formats similar to those used in the instructional materials of one program and not another. Closely related to this is the issue of "teaching for the test"—when this implies shaping the program to fit incidental features of a test such as item formats. We may as well throw in here the issue of test-wiseness itself as a program outcome—that is, the teaching of behaviors which, whether intended to do so or not, help children perform well on tests—since it bears on the overall problem of ways in which a program might achieve superior test scores without any accompanying superiority in actual learning of content. In short, children in some programs might simply get better at taking tests.

If one looks at the Direct Instruction and Behavior Analysis models, with their emphasis on detailed objectives and close monitoring of student progress, and compares them to EDC Open Education, with its disavowal of performance objectives and repudiation of standardized testing, it is tempting to conclude in the absence of any evidence that the former models must surely have turned out children better prepared to look good on tests, regardless of the children's true states of competence. Without wishing to prejudge the issue, we must emphasize that it is an empirical question to what extent children schooled in the various Follow Through models were favored or disfavored with respect to the process of testing itself.

In general, children involved in the Follow Through evaluation were subjected to more standardized testing than is normal. Since studies of test-wiseness indicate rapidly diminishing returns from increasing amounts of familiarization with testing (Cronbach, 1960), there is presumptive evidence against claims that differential amounts of test-taking among models could be significant in accounting for test-score differences. It should be possible to investigate this matter with Follow Through data, though not from the published data. Children in the final Follow Through evaluation had been subjected to from two to five rounds of standardized testing. Accordingly it should be possible to evaluate the effect of frequency of previous testing on third-grade test scores.

There are, however, numerous ways in which Follow Through experience could affect children's behavior during testing. The amount of experience that children in any program had with actual test-taking is probably trivial in comparison to the amount of experience some children got in doing workbook pages and similar sorts of paper-and-pencil activi-

ties. And the nature of these activities might have varied from ones calling for constructed responses, quite unlike those on a multiple-choice test, to ones that amounted virtually to a daily round of multiple-choice test-taking. Programs vary not only in the amount of evaluation to which children are subjected but also in the manner of evaluation—be it covert, which might have little effect on the children, or face-to-face and oral, or carried out through group testing. Finally, given that testing conditions in the Follow Through evaluation were not ideal, it is probably relevant how well children in the various programs learned to cheat effectively—that is, to copy from the right neighbor.

Some or most of these variables could be extracted from available information, and it would be then possible to carry out analyses showing the extent to which they account for test scores and for the score differences between models. Only through such a multivariate empirical investigation could we hope to judge how seriously to take suggestions that the score differences among models were artificial. Until that time, insinuations about “teaching for the test” must be regarded as mere prejudice.

What Do The Results Mean?

What we have tried to establish so far is that there are significant achievement test differences between Follow Through models and that, so far as we can tell at present, these test score differences reflect actual differences in school learning. Beyond this point, conclusions are highly conjectural. Although our main purpose in this paper has been simply to clarify the empirical results of the Follow Through experiment, we shall venture some interpretive comments, if for no other purpose than to forestall possible misinterpretations.

The two high-scoring models according to our analysis are Direct Instruction and Behavior Analysis; the two low-scoring are EDC Open Education and Responsive Education. If there is some clear meaning to the Follow Through results, it ought to emerge from a comparison of these two pairs of models. On the one hand, distinctive characteristics of the first pair are easy to name: sponsors of both the Direct Instruction and Behavior Analysis models call their approaches “behavioral” and “structured” and both give a high priority to the three R’s. EDC and Responsive Education, on the other hand, are avowedly “child-centered.” Although most other Follow Through models could also claim to be child-centered, these two are perhaps the most militantly so and most opposed to what Direct Instruction and Behavior Analysis stand for.

We can infer from Follow Through results that getting teachers of disadvantaged children to use more direct instructional methods as opposed to more informal ones will lead to superior achievement in commonly tested basic skills.

Thus we have, if we wish it, a battle of the philosophies, with the child-centered philosophy coming out the loser on measured achievement, as it has in a number of other experiments (Bennett, 1976; Stallings, 1975; Bell and Switzer, 1973; Bell, Zipousky & Switzer, 1976). This is interesting if one is keen on ideology, but it is not very instructive if one is interested in improving an educational program. Philosophies don’t teach kids. Events teach kids, and it would be instructive to know what kinds of events make the difference in scholastic achievement that we have observed.

The teaching behavior studies of Brophy & Good (1974), Rosenshine (1976), and Stallings & Kaskowitz (1974) are helpful on this point. Generally they contrast direct with informal teaching styles, a contrast appropriate to the two kinds of models we are comparing. Consistently it is the more direct methods, involving clear specifications of objectives, clear explanations, clear corrections of wrong responses, and a great deal of “time on task,” that are associated with superior achievement test performance. The effects tend to be strongest with disadvantaged children.

These findings from teacher observation studies are sufficiently strong and consistent that we may reasonably ask what if anything Follow Through results add to them. They add one very important element, the element of experimental change. The teacher observation studies are correlational. They show that teachers who do *x* get better achievement results than those who do *y*. The implication is that if the latter teachers switched from doing *y* to doing *x*, they would get better results, too; but correlational studies can’t demonstrate that. Perhaps teachers whose natural inclination is to do *y* will get worse results if they try to do *x*. Or maybe teachers who do *y* can’t or worse won’t do *x*. Or maybe *x* and *y* don’t even matter; they only serve as markers for unobserved factors that really make the difference.

The Follow Through experiment serves, albeit imperfectly, to resolve these uncertainties. Substantial resources were lavished on seeing to it that

teachers didn't just happen to use direct or informal methods according to their inclinations but rather that they used them according to the intent of the model sponsors. The experimental control was imperfect because communities could choose what Follow Through model to adopt, and in some cases, we understand, teachers could volunteer to participate. Nevertheless, it seems safe to assume that there was some sponsor effect on teacher behavior in all instances, so that some teachers who would naturally do *x* were induced to do *y* and vice-versa. Thus, with tentativeness, we can infer from Follow Through results that getting teachers of disadvantaged children to use more direct instructional methods as opposed to more informal ones will lead to superior achievement in commonly tested basic skills.

Before concluding, however, that what accounts for the superior achievement test scores of Direct Instruction and Behavior Analysis sites is their use of direct teaching methods, we should consider a more profound way in which these two models are distinguished from the others. These models are distinctive not only at the level of immediately observable teacher behavior but also at a higher level which may be called the systemic. One may observe a lesson in which the teacher manifests all the usual signs of direct teaching—lively manner, clear focus on instructional objectives, frequent eliciting of response from students, etc. One may return weeks later to find the same teacher with the same class manifesting the same direct teaching behavior—and still teaching the same lesson! The fault here is at the systemic level: the teacher is carrying out sorts of activities that should result in learning but is failing to organize and regulate them in such a way as to converge on the intended objectives.

More effective teachers—and this includes the great majority—function according to a convergent system. Consider a bumbling Mr. Chips introducing his pupils to multiplication by a two-digit multiplier. He demonstrates the procedure at the chalkboard and then discovers that most of the students cannot follow the procedure because they have forgotten or never learned their multiplication facts. So he backs up and reviews these facts, then demonstrates the algorithm again and assigns some practice problems. Performance is miserable, so he teaches the lesson again. By this time some children get it, and they teach others. With a bit of help, most of the class catches on. Mr. Chips then gives special tutoring, perhaps with use of supplementary concrete materials, to the handful of students who haven't yet got it. Finally everyone has learned the multiplication algorithm except for the slowest pupils in the class—who, as a matter of fact, haven't yet

learned to add either.

Although none of the procedures used by Mr. Chips are very efficient, he applies them in a convergent way so that eventually almost all the children reach the instructional objective. Some of his procedures may not have a convergent effect at all. For instance, he may assign practice worksheets to pupils who haven't yet grasped the algorithm, and the result is that they merely practice their mistakes (a divergent activity). But the overall effect is convergent. Given more efficient activities, convergence on the instructional goal might be more rapid and it might include the pupils who fail at the hands of Mr. Chips. But the difference in effectiveness, averaged over all pupils, would probably not be great. This

What Direct Instruction and Behavior Analysis provide are more fully developed instructional systems than teachers normally employ. They provide more systematic ways of determining whether children have the prerequisite skills before a new step in learning is undertaken, more precise ways of monitoring what each child is learning or failing to learn, and more sophisticated instructional moves for dealing with children's learning needs.

convergent property of teaching no doubt contributes, as Stephens (1967) has suggested, to the scarcity of significant differences between teaching methods. Unless severely constrained, most teachers will see to it that, one way or another, their students reach certain goals by the end of the term.

We suggest that teaching performance of the kind just described be taken as baseline and that innovative educational practices, such as those promoted by the Follow Through sponsors, be judged in relation to that baseline. What would happen to the teaching of our Mr. Chips if he came under the supervision of a Follow Through sponsor? It seems fairly clear that his system for getting students to reach certain goals by the end of the term would be enhanced if he took guidance from a Direct Instruction or Behavior Analysis sponsor but that it might well be disrupted by guidance from one of the more child-centered sponsors.

What Direct Instruction and Behavior Analysis provide are more fully developed instructional systems than teachers normally employ. They provide

more systematic ways of determining whether children have the prerequisite skills before a new step in learning is undertaken, more precise ways of monitoring what each child is learning or failing to learn, and more sophisticated instructional moves for dealing with children's learning needs. Open Education and Responsive Education, on the other hand, because of their avowed opposition to making normative comparisons of students or thinking in terms of deficits, will tend to discourage those activities whereby teachers normally discover when children are not adequately prepared for a new step in learning or when a child has mislearned or failed to learn something. Also, because of their preference for indirect learning activities, these models will tend to make teaching less sharply focused on achieving specific earnings and remedying specific lacks.

Open Education and Responsive Education, because of their avowed opposition to making normative comparisons of students or thinking in terms of deficits, will tend to discourage those activities whereby teachers normally discover when children are not adequately prepared for a new step in learning or when a child has mislearned or failed to learn something.

Of course, child-centered educators will wish to describe the matter differently, arguing that they do have a well-developed system for promoting learning; but it is a different kind of system pursuing different kinds of goals from those pursued by the direct instructional approaches. They will point out that child-centered teachers devote a great deal of effort to identifying individual pupils' learning needs and to providing learning experiences to meet these needs; it is just that their efforts are more informal and intuitive, less programmed. Child-centered education, they will argue, is different, not inferior.

One is inclined automatically to assent to this live-and-let-live assessment, which relegates the differences between educational methods to the realm of personal values and ideology. But surely the Follow Through experiment and any comparative evaluation will have been in vain if we take this easy way out of the dilemma of educating disadvantaged children.

This easy way of avoiding confrontation between the two approaches can be opposed on both empirical and theoretical grounds. Empirically, child-centered approaches have been unable to demonstrate any off-setting advantages to compensate for their poor showing in teaching the three R's. House et al (1978a) have argued that the selection of measures used in the Follow Through evaluation did not give child-centered approaches adequate opportunity to demonstrate their effects. This may be true to a degree, but it is certainly not true that child-centered approaches had no opportunity to demonstrate effects relevant to their purposes. One had better not be a perfectionist when it comes to educational evaluation. No measure is perfectly correlated to one's objectives. The most one can hope for is a substantial correlation between obtained scores on the actual measures and true scores on the ideally appropriate measures that one wishes existed but do not.

When child-centered educators purport to increase the self-esteem of disadvantaged children and yet fail to show evidence of this on the Coopersmith Self-Concept Inventory, we may ask what real and substantial changes in self-esteem would one expect to occur that would not be reflected in changes on the Coopersmith? Similarly for reasoning and problem-solving. If no evidence of effect shows on a test of non-verbal reasoning, or a reading comprehension test loaded with inferential questions, or on a mathematical problem solving test, we must ask why not? What kinds of real, fundamental improvements in logical reasoning abilities would fail to be reflected in any of these tests?

If these remarks are harsh, it is only because we believe that the question of how best to educate disadvantaged children is sufficiently serious that a policy of live-and-let-live needs to be replaced by a policy of put-up-or-shut-up. Certainly the cause of educational betterment is not advanced by continual appeal to nonexistent measures having zero or negative correlations with existing instruments purporting to measure the same thing. Among the numerous faults that we have found with the House committee's report, their use of this appeal is the only one that deserves the label of sophistry.

Critique of the Child-Centered Approach

What follows is an attempt at a constructive assessment of the child-centered approach as embodied in the Open Education and Responsive Education models. By constructive we mean that we take seriously the goals of these models and that our

If these remarks are harsh, it is only because we believe that the question of how best to educate disadvantaged children is sufficiently serious that a policy of live-and-let-live needs to be replaced by a policy of put-up-or-shut-up.

interest is in realizing the goals rather than in scraping them in favor of others. These remarks are by way of preface to the following observation: *child-centered approaches have evolved sophisticated ways of managing informal educational activities but they have remained at a primitive level in the design of means to achieve learning objectives.*

We are here distinguishing between two levels at which a system of teaching may be examined. At the *management* level, an open classroom and a classroom running according to a token economy, for example, are radically different, and while there is much to dispute in comparing them, it is at least clear that both represent highly evolved systems. When we consider the *instructional design* level, however, the difference is more one-sided. Child-centered approaches rely almost exclusively on a form of instruction that instructionally-oriented approaches use only when nothing better can be found.

This primitive form of instruction may be called *relevant activity*. Relevant activity is what teachers must resort to when there is no available way to teach children how to do something, no set of learning activities that clearly converge on an objective. This is the case, for instance, with reading comprehension. Although there are some promising beginnings, there is as yet no adequate "how-to-do-it" scheme for reading comprehension. Accordingly, the best that can be done is to engage students in activities relevant to reading comprehension—for instance, reading selections and answering questions about the selections. Such activities are relevant in that they entail reading comprehension, but they cannot be said to teach reading comprehension.

For many other areas of instruction, however, more sophisticated means have been developed. There are, for instance, ways of teaching children *how* to decode in reading and *how* to handle equalities and inequalities in arithmetic (Engelmann, Ref. Note 2). The instructional approaches used in Direct Instruction and Behavior Analysis reflect years of analysis and experimentation devoted to finding ways of going beyond relevant activity to forms of

instruction that get more directly at cognitive skills and strategies. This effort has been successful in some areas, not so successful in others, but the effort goes on. Meanwhile, child-centered approaches have tended to fixate on the primitive relevant activities form of instruction for all their instructional objectives.

The contrast of sophistication in management and naiveté in instruction is visible in any well-run open classroom. The behavior that meets the eye is instantly appealing—children quietly absorbed in planning, studying, experimenting, making things—and one has to marvel at the skill and planning that have achieved such a blend of freedom and order. But look at the learning activities themselves and one sees a hodge-podge of the promising and the pointless, of the excessively repetitious and the excessively varied, of tasks that require more thinking than the children are capable of and tasks that have been cleverly designed to require no mental effort at all (like exercise sheets in which all the problems on the page have the same answer). The scatteredness is often appalling. There is a little bit of phonics here and a little bit of phonics there, but never a sufficiently coherent sequence to enable a kid to learn

Child-centered educators have evolved a style of school life that has much in its favor. Until they develop an effective pedagogy to go with it, however, it does not appear to be an acceptable way of teaching disadvantaged children.

how to use this valuable tool. Materials have been chosen for sensorial appeal or suitability to the system of management. There is a predilection for cute ideas. The conceptual analysis of learning problems tends to be vague and irrelevant, big on name-dropping and low on incisiveness.

There does not appear to be any intrinsic reason why child-centered educators should have to remain committed to primitive instructional approaches. So far, child-centered educators have been able to gain reassurance from the fact that for the objectives they emphasize—objectives in comprehension, thinking, and feeling—their approaches are no more ineffective than anyone else's. But even this defense may be crumbling. Instructional designers, having achieved what appears to be substantial success in improving the teaching of decoding in reading, basic mathematical concepts and

operations, spelling, and written English syntax, are now turning more of their attention to the kinds of goals emphasized by child-centered educators. Unless thinkers and experimenters committed to child-centered education become more sophisticated about instruction and start devoting more attention to designing learning activities that actually converge on objectives, they are in danger of becoming completely discredited. That would be too bad. Child-centered educators have evolved a style of school life that has much in its favor. Until they develop an effective pedagogy to go with it, however, it does not appear to be an acceptable way of teaching disadvantaged children.

Notes

1. Reduced analyses were performed, dropping TEEM and Cognitive Curriculum from the analysis. These were the two most unstable models in the sense of shifting most in relative performance depending on the choice of covariates. Moreover, Cognitive Curriculum had deviant relations between criteria and covariates, showing for instance negative relationships between achievement and SES. The only effect of removing these models, however, was to increase the number of significant differences between the two top scoring models and the other models.
2. Examined closely, the House et al statement is a bit slippery. Since the MAT is a norm-referenced, (not a criterion-referenced) test, it is of course "reckless" to infer any particular attainments at all from test scores. All we know is how a person or group performs in comparison to others. If, for example, the criterion for "ability to write a story" is set high enough, it would be reckless to suppose that any third-grader had attained it.
3. The obvious targets for the charge of emphasizing rote learning are Direct Instruction and Behavior Analysis. However, the Direct Instruction sponsors explicitly reject rote memorization (Bock, Stebbins, & Proper, 1977, p. 65) and the Behavior Analysis model description makes no mention of it. House, Glass, McLean, and Walker seem to have fallen into the common fallacy here of equating direct instruction with rote learning. If they are like most university professors, they probably rely extensively on direct instruction themselves and yet would be offended by the suggestion that this means they teach by rote.

Reference Notes

1. Becker, W.C., & Carnine, D.W. Direct Instruction—A behavior-based model for comprehensive educational intervention with the disadvantaged. Paper presented at the VIII Symposium on Behavior Modification, Caracas, Venezuela, February, 1978. Division of Teacher Education, University of Oregon, Eugene, Oregon.
2. Engelmann, S. *Direct Instruction*. Seminar presentation. AERA, Toronto, March, 1978.

References

- Bell, A.E., & Switzer, F. (1973). Factors related to pre-school prediction of academic achievement: Beginning reading in open area vs. traditional classroom systems. *Manitoba Journal of Education*, 8, 22-27.
- Bell, A.E., Zipuvsky, M.A., and Switzer, F. (1977). Informal or open-area education in relation to achievement and personality. *British Journal of Educational Psychology*, 46, 235-243.
- Bennett, N. (1976). *Teaching styles and pupil progress*. Cambridge, Mass.: Harvard University Press.
- Brophy, J.E., & Good, T.L. (1974). *Teacher-student relationships: Causes and consequences*. New York: Holt, Rinehart & Winston.
- Cronbach, L.J. (1960). *Essentials of psychological testing*. (2nd ed.). New York: Harper & Brothers.
- House, E.R., Glass, G.V., McLean, L.F., and Walker, D.F. (1978a). No Simple Answer: Critique of the "Follow Through" evaluation. *Harvard Educational Review*, 28(2), 128-160.
- House, E.R., Glass, G.V., McLean, L.F., and Walker, D.F. (1978b). Critiquing a Follow Through evaluation. *Phi Delta Kappan*, 59(7), 473-474.
- Rosenshine, B. Classroom Instruction. (1976). In *Seventy-fifth Yearbook of the National Society for the Study of Education* (Part 1). Chicago: University of Chicago Press.
- St. Pierre, R.G., Anderson, R.B., Proper, E.C., and Stebbins, L.B. (1978). That Follow Through evaluation. *Phi Delta Kappan*, 59(10), 729.
- Stallings, J.A., & Kaskowitz, D.H. (1974). *Follow Through classroom observation evaluation-1972-1973*. Menlo Park, Cal.: Stanford Research Institute.
- Stallings, J. (1975). Implementation and child effects of teaching practices in Follow Through classrooms. *Monographs of the Society for Research in Child Development*, 40(7-8, Serial No. 163).
- Stebbins, L.B., St. Pierre, R.G., Proper, E.C., Anderson, R.B., and Cerva, T.R. (1977). *A planned variation model. Vol. IV-A Effects of Follow Through models*. U.S. Office of Education.
- Stephens, J. (1967). *The process of schooling*. New York: Holt, Rinehart & Winston.

Sponsor Findings from Project Follow Through

Wesley C. Becker and Siegfried Engelmann
University of Oregon

The final report of the National Evaluation of Project Follow Through, a comparative study of different approaches to teaching economically disadvantaged children in the primary grades, shows that the Direct Instruction Model (University of Oregon) was largely successful in assisting disadvantaged children in catching up with their middle-class peers in academic skills. This demonstration is the first to show that compensatory education can work.

The Direct Instruction Model emphasizes small-group face-to-face instruction by teachers and aides using carefully sequenced lessons in reading, arithmetic, and language. These programs were designed by Siegfried Engelmann using modern behavioral principles and advanced programming strategies (Becker, Engelmann, & Thomas, 1975), and are published by Science Research Associates under the trade name DISTAR. The program directors, Professor Wesley C. Becker and Siegfried Engelmann, attribute its success to the technological details, the highly specific teacher training, and careful monitoring of student progress. The closest rival to the Direct Instruction Model in overall effects was another behaviorally-based program, the University of Kansas Behavior Analysis Model. Child-centered, cognitively focused, and open classroom approaches tended to perform poorly on all measures of academic progress.

Design

The National Evaluation of Follow Through used a planned variation design to provide a broad-range comparison of educational alternatives for teaching the disadvantaged and find out "what works." Different models of instruction were tested in 139 communities and evaluated for stability of results over successive program years. Model programs were implemented in kindergarten through third grade. The descriptions of the nine major models in the National Evaluation are taken from the Abt Associates descriptions of the report.

The Open Classroom Model (Education Development Center, EDC) which is based on the British Infant School model;

Cognitively-Oriented Curriculum Model (High/Scope Educational Research Foundation) which is based on Piaget's theories;

The Responsive Education Model (Far West Laboratory for Educational Research) which is based on Glen Nimniet's work in structuring a teaching environment and uses a variety of techniques;

The final report of the National Evaluation of Project Follow Through shows that the Direct Instruction Model (University of Oregon) was largely successful in assisting disadvantaged children in catching up with their middle-class peers in academic skills. This demonstration is the first to show that compensatory education can work.

Bank Street Early Childhood Education Model (Bank Street College of Education) which is concerned with the development of the whole child.

Tucson Early Education Model (TEEM, University of Arizona) which is based on the language-experience approach of Marie Hughes that initially focused on teaching bilingual children; i

Language Development (Bilingual) Model (Southwest Educational Development Laboratory—SEDL) which utilizes programmed curricula for bilingual children (and others) focusing on language development;

Behavior Analysis Model (University of Kansas) which used modern principles of reinforcement and systematic classroom management procedures; and

Direct Instruction Model (University of Oregon).

For each sponsor, children were followed from entry to third grade in 4 to 8 kindergarten-entering sites, and some first-grade-entering sites. Comparison groups were also tested. The evaluation, referred to as "the largest controlled education experiment ever," included measures of Basic Skills, Cognitive Skills, and Affect.

Basic Skills were based on four subtests of the Metropolitan Achievement Test (MAT)—Word Knowledge, Spelling, Language, and Math Computation.

The Cognitive Skills included MAT Reading, MAT Math Concepts, MAT Math Problem Solving, and the Raven's Coloured Progressive Matrices.

The Affective Measures consisted of the Coopersmith Self-Esteem Inventory and the Intellectual Achievement Responsibility Scale (IARS). The Coopersmith measures children's feelings about themselves and school; the IARS measures the degree to which children take responsibility for their successes and failures.

Results

Adjusted Outcomes

Abt Associates used covariance analysis to adjust third-grade scores according to entry differences between experimental and comparison groups. An adjusted difference was defined as *educationally significant* if the difference between experimental and comparison group was at least one-fourth standard deviation unit in magnitude. This convention was adopted because when dealing with large groups, statistical significance can be very misleading.

Figures 1 to 3 show the performance of the various sponsors on the adjusted outcomes in comparison to control groups. An Index of Significant outcomes (ISO's) is used to show relative effects across models. ISO's are derived by taking the number of educationally significant **minus** outcomes for a sponsor and subtracting this from the number of significant **plus** outcomes.¹ This number (which may be negative) is divided by the total number of comparisons for a model and multiplied by 1000 to get rid of decimal points. The result is a number either positive or negative, that shows both the plus-minus direction and the consistency of each model's effects. If the direction is positive, it means that the model outperforms the controls. The larger the

number, the more consistently the model performs. If the number is negative, the control groups outperform the model.

Figure 1 compares the performance of different models on Basic Skills. Only three models achieve positive ISO's. The Direct Instruction Model is more than 270 ISO points above the nearest comparison (Florida). The Direct Instruction Model is about 700 points above the lowest program, EDC's Open Education Model.

Figure 2 compares models on *academic* Cognitive-Conceptual Skills.² Only two models have positive outcomes, and again the Direct Instruction Model is in first place (this time by over 225 ISO points above the second-place finisher and by over 800 ISO points above the lowest program [EDC]). The performance on Cognitive-Conceptual Skills demonstrates that those programs based on cognitive "theories" do not have the technological know-how to achieve positive results with poverty children, and that behaviorally-based Behavior Analysis Model also lacks the technology to teach Cognitive-Conceptual Skills.

Figure 3 compares models on Affective Measures. The Direct Instruction Model achieves the highest positive effect. Behavioral Analysis, Parent Education, and SEDL also have positive ISO's. Note that only those models that achieved positive effects on Basic Skills or Cognitive-Conceptual skills produce positive outcomes on Affective Measures. Note also that the cognitively-oriented programs (with the exception of Parent Education) perform as poorly on Affective Measures as they do on Academic Achievement. The high correlation between academic and affective outcomes suggest a need to re-evaluate some interpretations of what turns kids on and how they learn to feel good about themselves in school.

Grade-Equivalent and Percentile Performance

The Abt IV Report provides performance level data for four MAT measures: Total Reading, Total Math, Spelling, and Language. Tables 1-4 display percentiles on a one-fourth standard deviation scale. With this display, differences between sponsors of a quarter-standard deviation (e.g., an educationally significant difference) are easily detected, while the percentiles provide the "norm reference." The baseline at the 20th percentile represents average expectation for disadvantaged children.

Total Reading (Table 1). The Direct Instruction Model, the only one to show achievement above 3.0 grade level, is about one-half standard deviation above the mean of all other sponsors. It is nearly a quarter-standard deviation above the second-place model, Behavior Analysis.

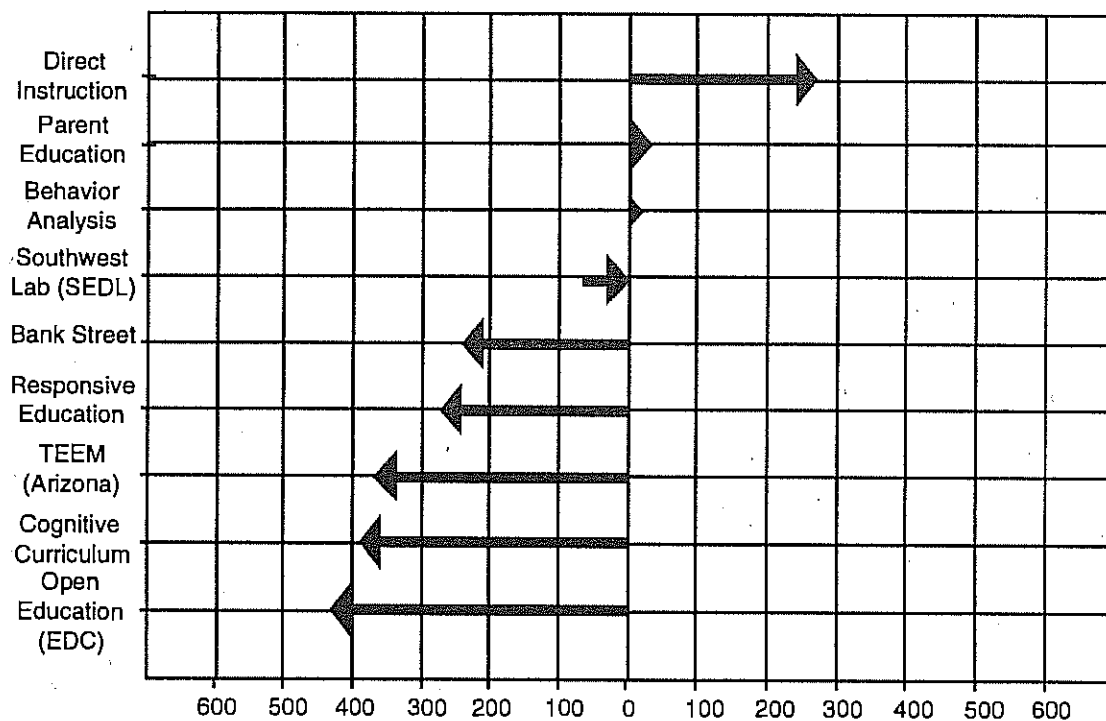


Figure 1

ISO's for basics measures (word knowledge, spelling, language, math computation)

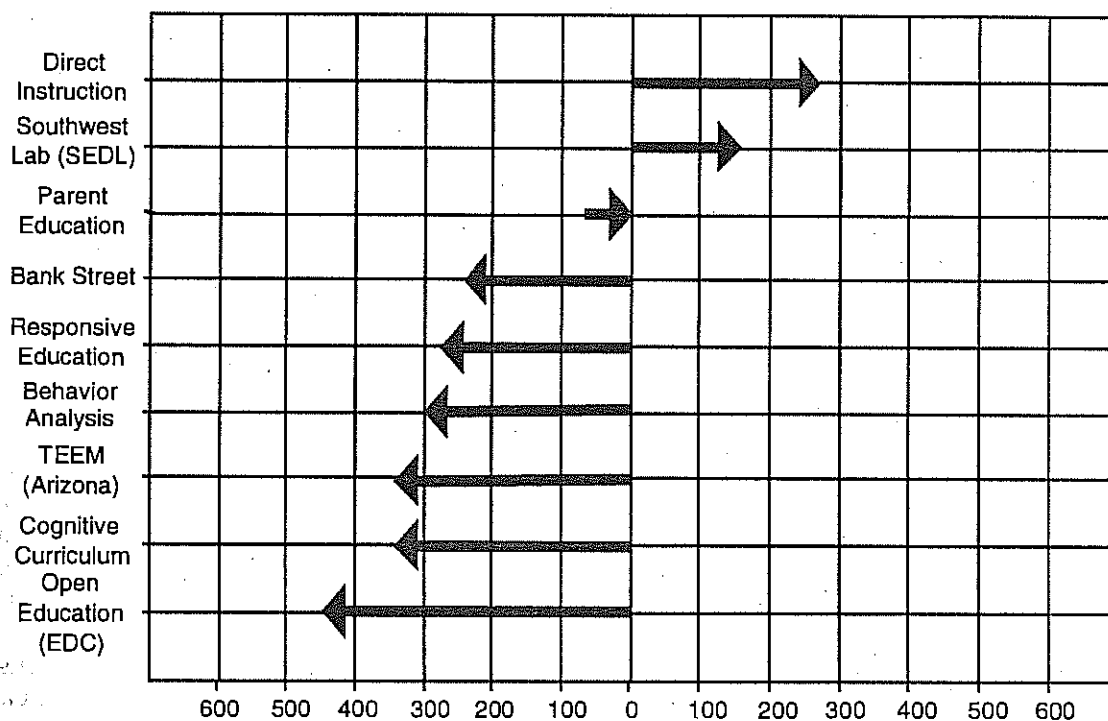


Figure 2

ISO's for cognitive measures (reading comprehension, math concepts, math problem solving)

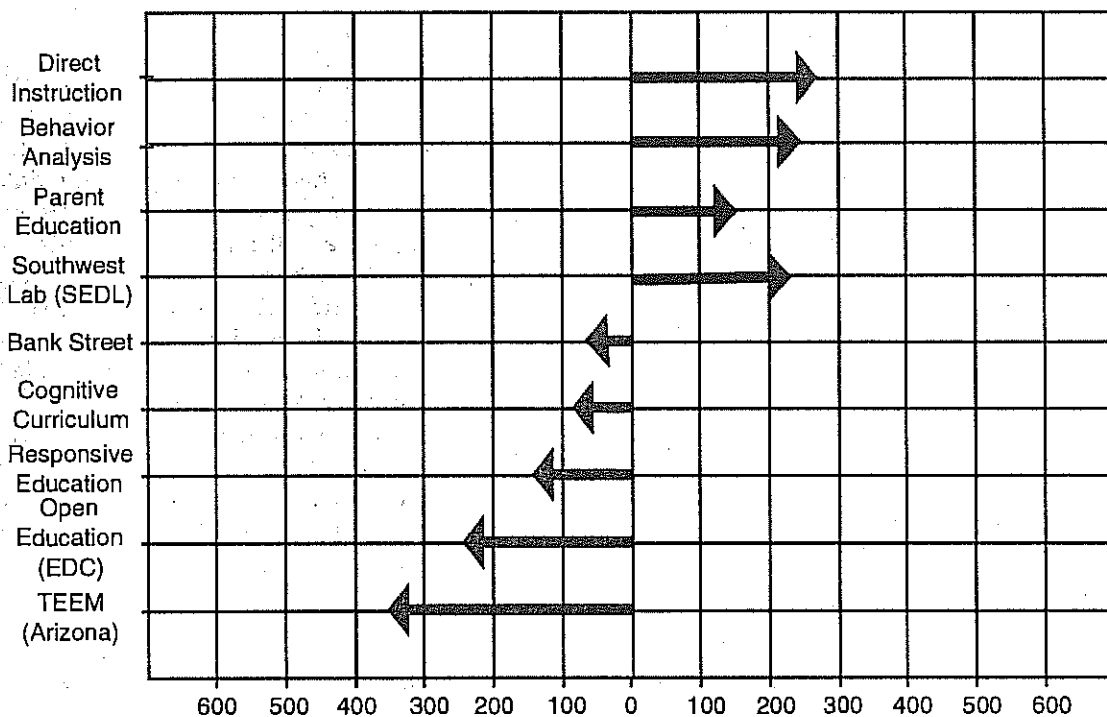


Figure 3

ISO's for affective measures (cooperation, self-esteem, intellectual achievement, responsibility scale)

Total Math (Table 2). The mean grade-equivalent scores for the Direct Instruction Model is 3.71, which is 48th percentile, and one full standard deviation about the average of all other sponsors. The Model is one-half standard deviation about the second-place model, again, Behavior Analysis.

Spelling (Table 3). The Direct Instruction Model achieves the 51 percentile and again leads all sponsors. The Behavior Analysis model, however, is a close second (49th percentile).

Language (Table 4). The Direct Instruction Model performs at the 4.0 grade level, or 50th percentile. It is three-fourths standard deviation above all other models. (No other model scores within one year of the Direct Instruction Model on grade-equivalent score.)

Sponsor Findings

Sponsor-collected data further support the above conclusions:

- A greater measurable and educationally significant benefit is present at the end of third grade for those who begin Direct Instruction in kindergarten than for those who begin in first grade (Becker and Engelmann, 1978; Gersten, Darch & Gleason, xx).

- Significant gains in IQ are found, which are largely maintained through third grade. Students entering the program with IQ's over 111 do not lose during the Follow Through years, though one might expect some repeated regression phenomena. The low-IQ children, on the other hand, display appreciable gains, even after the entry IQ has been corrected for regression artifact. Students with IQ's below 71 gain 17 points in the entering kindergarten sample and 9.4 points in the entering first-grade sample; gains for the children entering with IQ's in the 71-90 range are 15.6 and 9.2, respectively (Gersten, Becker, Heiry & White, 1984).

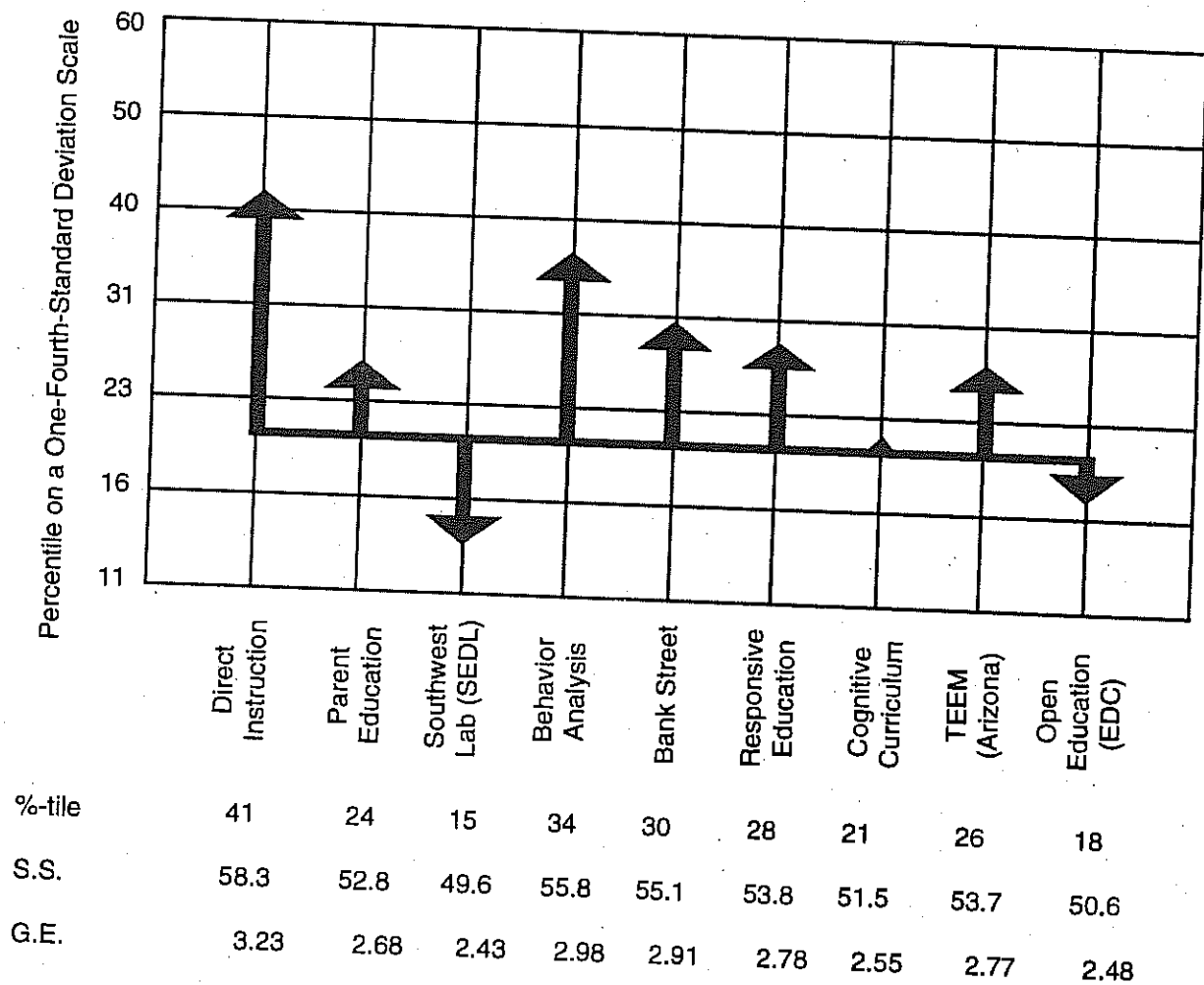
- Studies of low-IQ students (under 80) show the program is clearly effective with students who have a higher probability of failure. As indicated in Figures 3 and 4, these students gain nearly as much each year in reading (decoding) and math, as the rest of our students with higher IQ's—more than a year-per-year on the WRAT (Wide

Range Achievement Test) Reading and a year-per-year on MAT (Metropolitan Achievement Test) Total Math (Gersten et al., 1984).

- High school follow-up studies of Direct Instruction and comparison students were carried out in five districts. All the significant differences favored Direct Instruction students: five on academic measures, three on attendance, two on college acceptance and three on reduced retention rates (Gersten and Keating, 1987).
- The model generalizes across both time and across populations. The Department of Education has a Joint Dissemination Review Panel that validates educational

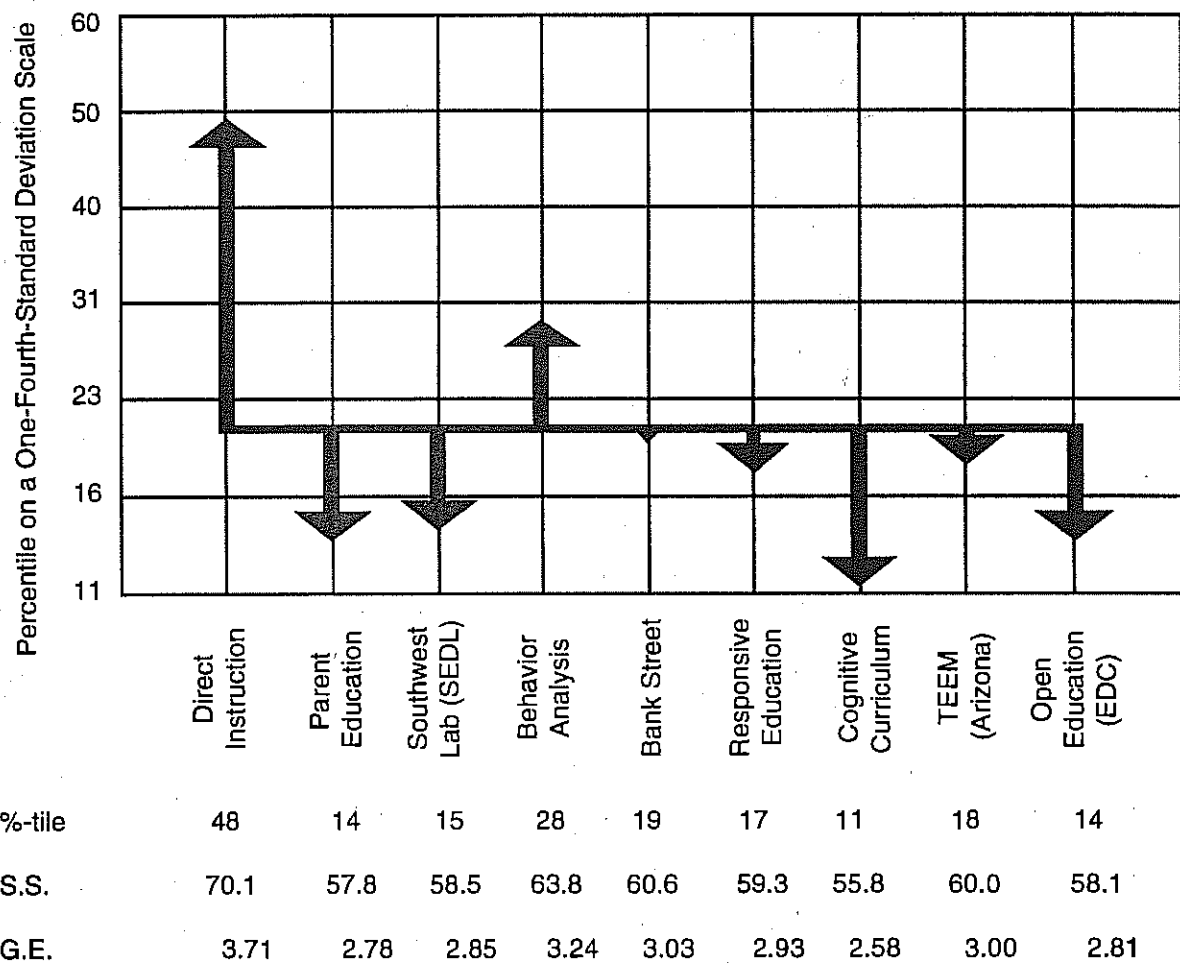
programs as exemplary and qualifies them for national dissemination. During the 1980-81 school year, the last of the 12 Direct Instruction Follow Through projects were submitted for validation. Of the 12 districts, 11 had 8 to 10 years of data on successive groups of children. The schools sampled a full range of students: large cities (New York; San Diego; Washington, D.C.); middle-sized cities (Flint, MI; Dayton, OH; E. St. Louis, IL); rural white communities (Flippin, AR; Smithville, TN); a rural Black community (Williamsburg, SC); Mexican American communities (Uvalde, TX; E. Las Vegas, NM); and an American Indian community (Cherokee, NC). One hundred percent of the projects were certified as exemplary in reading and

Table 1
M.A.T. Total Reading



Grade equivalent for 50th percentile is 3.5

Table 2
M.A.T. Total Math



Grade equivalent for 50th percentile is 3.75

mathematics for the primary grades, thus providing replication over 8 to 10 years and in a dozen quite diverse communities.

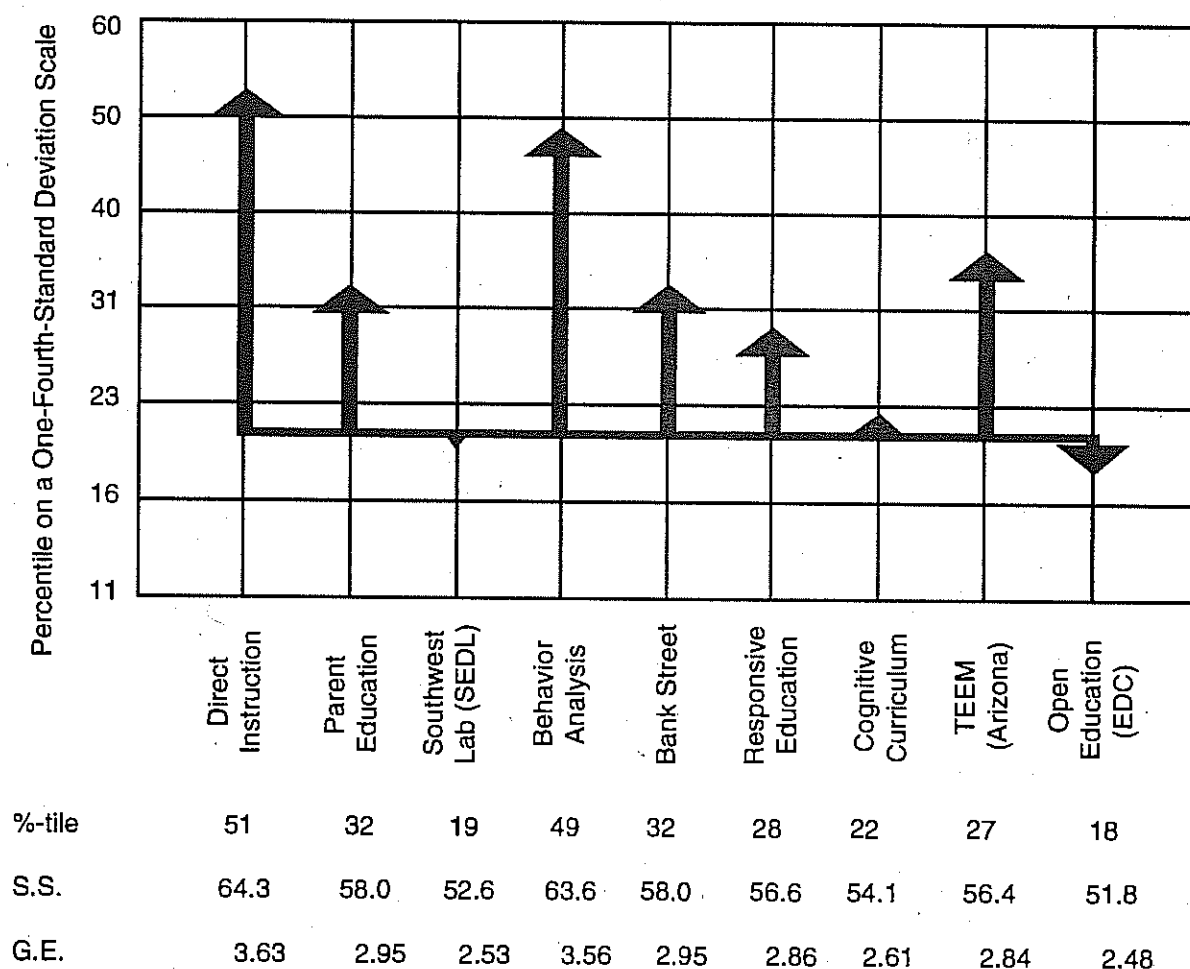
- Research on implementation found consistent high-to-moderate relationships between observed level of model implementation and classroom achievement gains in reading. At least for highly structured models of instruction, degree of implementation can be measured in a reliable and valid fashion (Gersten, Carnine, Zoref, Cronin, 1986).

Two conclusions seem of special interest, especially in view of the wave of programs recently initiated in major urban areas to improve the teaching of basic skills. The first is that teachers at first may react negatively to—or be confused by—inten-

sive, structured, in-class training (or technical assistance). Yet, ultimately at least half of the teachers found this to be one of the most positive features of the intervention.

The other key finding is that many teachers altered their reactions to structured educational models after they saw the effects of this program with their students on a day-to-day basis. Often this transformation took many months. At the beginning teachers were far from enthusiastic about the program and tended to feel that too much time was devoted to academics. Not enough was set aside for "fun" or creative activities. Yet their strong support by the end of the second year was unequivocal. From teacher interview data collected over two years, there can only be one main explanation for this, namely, the effect of the Direct Instruction Model on student performance. Time and again the teachers marveled at the new academic skills their pupils

Table 3
M.A.T. Spelling



Grade equivalent for 50th percentile is 3.6

demonstrated. Teachers reported anecdotal evidence of growth well before the standardized achievement tests were administered (Cronin, 1980).

Implications of the Direct Instruction Findings

The Follow Through data and our extensive experience in the field attempting to generate changes in school systems permit tentative answers to a number of major issues in the field today.

Will Money and Comprehensive Services Do the Job?

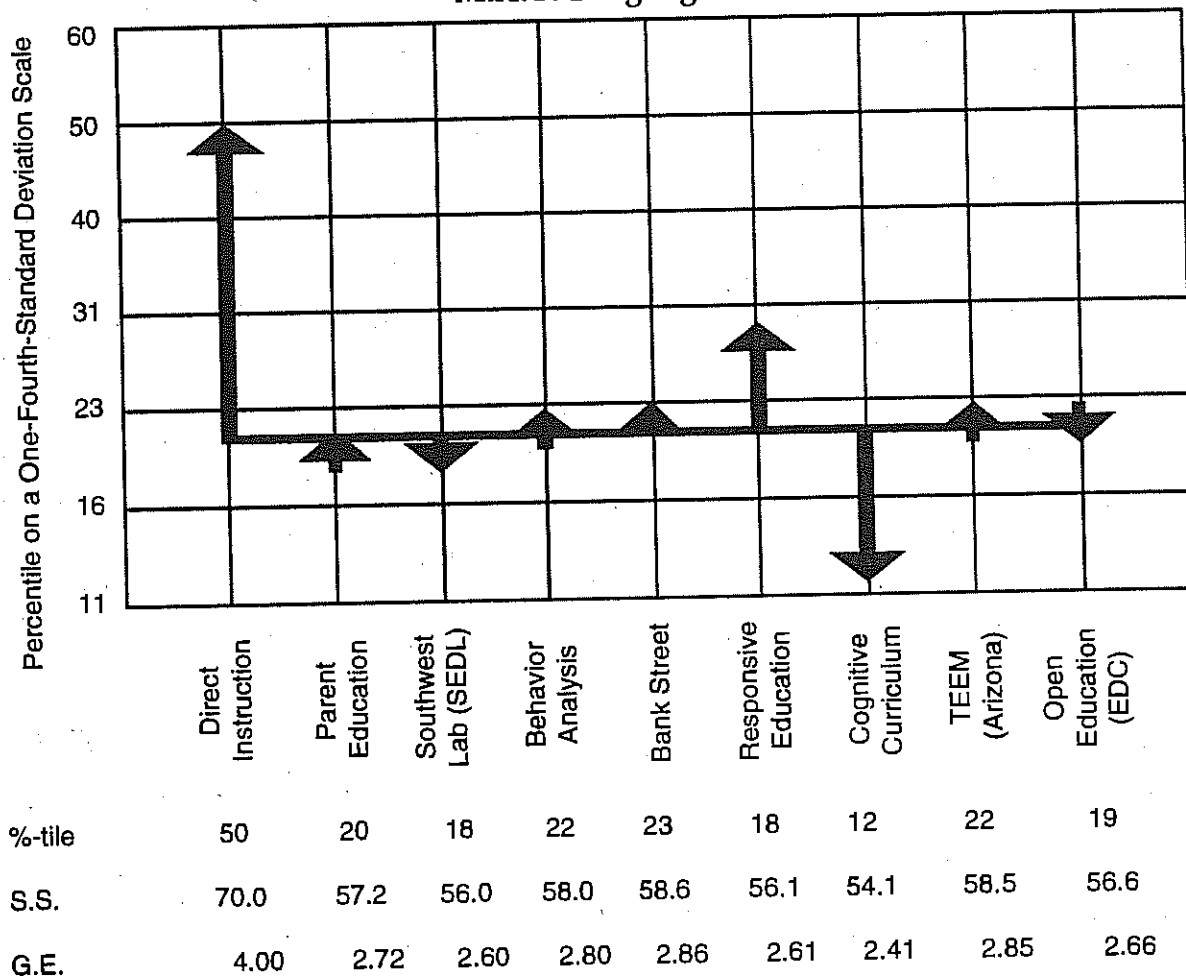
Each of the sponsors in Follow Through had about the same amount of money to provide comprehensive services and an educational program. Most sponsors had two aides in most classrooms, and spent about \$350 per child above basic school support on the educational component. The Abt data provide a convincing demonstration that money,

good will, people, material, Hawthorne effect, health programs, dental programs, and hot lunches do not cause gains in achievement. All Follow Through sponsors had these things, and most failed to do the job in basic instruction.

Does Individualization Require Many Approaches?

The programs that failed the most in terms of educational achievements were those oriented to individual needs in instruction. The popular belief that it is necessary to teach different students in different ways is, for the most part, a fiction. The requirements for sequencing an instructional program are determined by what is to be taught, not who. In the DISTAR programs used by the Direct Instruction Model, each child faces the same sequence of tasks and the same teaching strategies. What is individualized is entry level, when corrections are used, reinforcement procedures, and number of practice trials to mastery.

Table 4
M.A.T. Language



Grade equivalent for 50th percentile is 4.0

Is Self-Directed Learning Best?

A common assumption arising from dominant subjective education philosophies is that self-directed learning is the only meaningful learning. Direct Instruction is said to produce isolated rote learning, not "meaningful" learning. The Follow Through results obviously demonstrate such an assumption to be false. The students performing best on all measures of higher cognitive processes were from the Direct Instruction Model. The assumption about the value of self-directed learning probably arises from observing young children (as Piaget did) interacting with the physical environment. The physical environment directly reinforces and punishes different responses. However, there is no way a child can learn the arbitrary conventions of a language system without someone who knows that system providing systematic teaching (including modeling of appropriate language usage). In addition, there can be no question that smart adults can

organize and sequence experiences that will teach concepts and problem-solving skills better than children.

Why is Improvement in Reading Comprehension Hard to Achieve?

The Abt IV Report notes that successful outcomes were harder to come by in reading comprehension than in other skill areas. Only the Direct Instruction program made significant and sustained gains in this area. Even then, we only reached the 40th percentile on MAT Reading. Becker (1977) analyzed the Follow Through data and other data on reading, and concluded that schools are not designed to teach the English language to "poor kids" (e.g., to children whose parents, on the average, are less well-versed in knowledge of standard English). Schools are basically designed for white, middle-class children, and leave largely to parents the teaching of a most basic building block for intelligent behavior—

namely, words and their referents.

Why Do Economically Disadvantaged Students Continue to Do Poorly in School?

In general, economically disadvantaged students come to school with less knowledge relevant to succeeding in school. Thus, teaching these students requires teachers with different attitudes and skills, and more patience than is typically required. Colleges of education and schools are not organized or administered to develop and support teachers with these attributes. To coin a malapropism, "there is a way, but no will." Students from low-income families do not need to fail in schools. They can be taught.

In summary, through the careful design of curricula, classroom procedures, and training procedures, the DI Follow Through Model was able to achieve a major goal of compensatory education—improving the academic performance of economically disadvantaged children to (or near) median national levels. Only one other major model in the Follow Through experiment (the University of Kansas Behavior Analysis Model) came close to matching this achievement. The DI Model also performed best on measures of affective outcomes, such as self-esteem. Follow-up studies, through primary and secondary levels, show strong continuing effects in

terms of academic performance at the primary level, and better attendance, fewer grade retentions, and increased college acceptance at the high school level.

The Communities

The communities which have used the Direct Instruction Model are Providence, RI, Brooklyn, NY (P.S. 137), Washington D.C., Cherokee, NC, Williamsburg County, SC, Dayton, OH, E. St. Louis, IL, Flint, MI, Grand Rapids*, MI, West Iron County*, MI, Smithville, TN, Tupelo, MS, Racine*, WI, Todd County, SD, Rosebud Tribe, SD, Flippin, AR, Uvalde, TX, Dimmitt*, TX, E. Las Vegas, NM.

Notes

¹The Abt analysis provides the two comparisons for each measure. One with a local control group and the other with a pooled national control group. A comparison was counted plus if either comparison was plus, and minus if either was minus. Use of alternative decisions rules would not change the relative rankings of models.

²The Raven's Coloured Progressive Matrices result is not included with the data graphed because it is not an academic skill. Only 3 of 27 comparisons for all nine sponsors showed a positive outcome on the Raven's suggesting that this test does not reflect what was being taught by sponsors. Direct Instruction shows a negative ISO on this measure, but would still rank 1 if it were included.

*No longer in Follow Through.

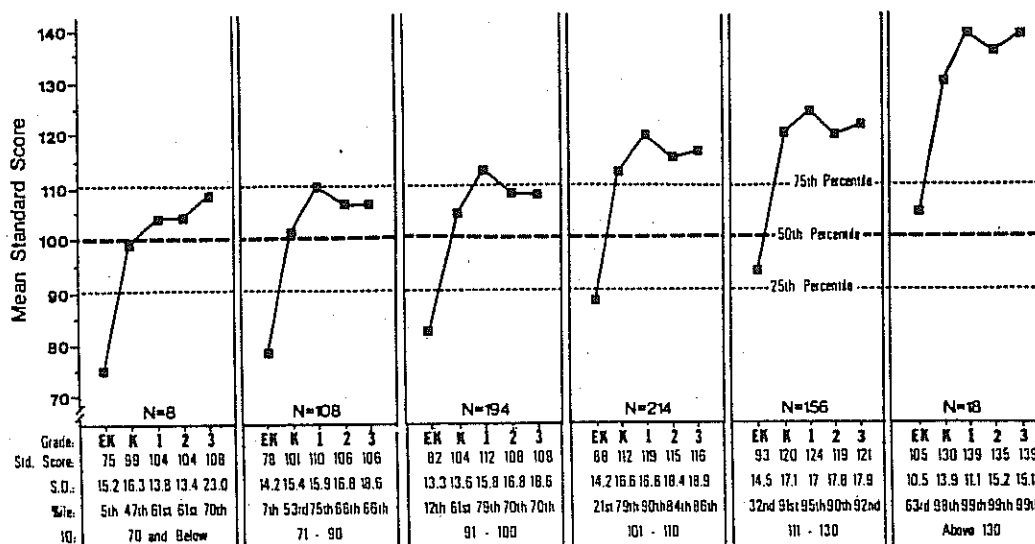


Figure 4

Yearly gains (K-3) on WRAT Reading (decoding) for students according to IQ blocks (K-starting students).

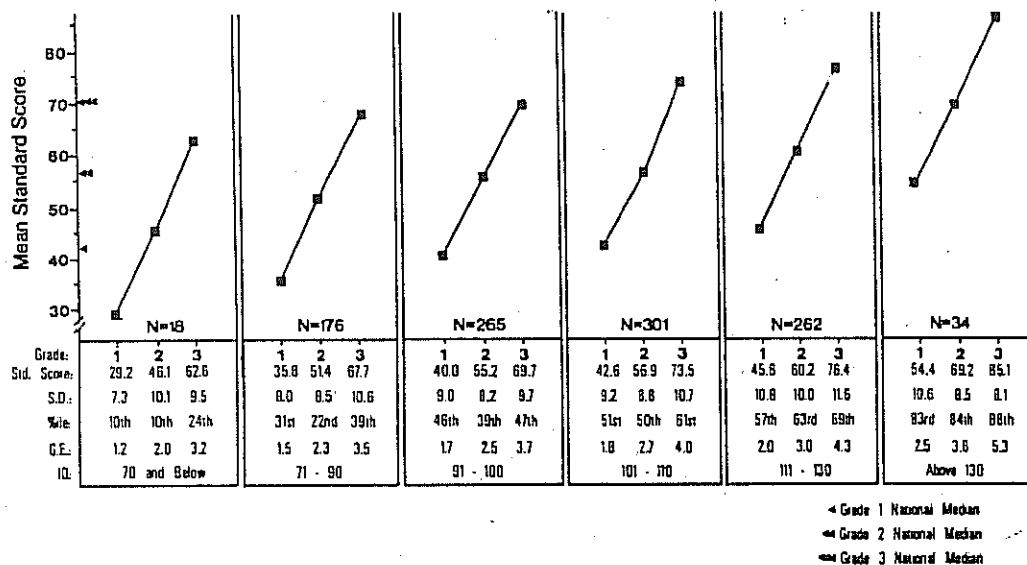


Figure 5

Yearly gains (1-3) on MAT Total Math for students according to IQ blocks (K-starting students).

References

- Abt Associates. (1977). *Education as experimentation: A planned variation model (Vol. IV)*. Cambridge, MA: Author.
- Becker, W.C., Engelmann, S., & Thomas, D.R. (1975). *Teaching 2: Cognitive Learning and Instruction*. Chicago: Science Research Associates.
- Becker, W.C., & Engelmann, S. (1976). *Analysis of achievement data on six cohorts of low income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model (Technical Report #76-1)*. Eugene, OR: University of Oregon, Office of Education, Follow Through Project.
- Becker, W., & Engelmann, S. (1978). *Analysis of achievement data on six cohorts of low income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model (Technical Report #78-1)*. Eugene, OR: University of Oregon, Office of Education, Follow Through Project.
- Bereiter, C. (1967). *Acceleration of intellectual development in early childhood*. Final Report Project No. 2129, Contract No. OE 4-10-008. Urbana, IL: University of Illinois, College of Education.
- Cronin, D. P. (1980). *Implementation study, year 2, Instructional staff interviews*. Los Altos, CA: Emrick.
- Engelmann, S. (1967). Teaching formal operations to preschool children. *Ontario Journal of Educational Research*, 9 (3), 193-207.
- Engelmann, S. (1968). The effectiveness of direct verbal instruction on IQ performance and achievement in reading and arithmetic. In J. Hellmuth (Ed.), *Disadvantaged Child*, Vol. 3. New York: Bruner Mazel.
- Gersten, R., Becker, W., Heiry, T., & White. (1984). Entry IQ and yearly academic growth in children in Direct Instruction programs: A longitudinal study of low SES children. *Educational Evaluation and Policy Analysis*, 6(2), 109-121.
- Gersten, R., & Keating, T. (1987). Improving high school performance of "at risk" students: A study of long-term benefits of direct instruction. *Educational Leadership*, 44(6), 28-31.
- Nero and Associates, Inc.. (1975). *Follow Through. A description of Follow Through sponsor implementation processes*. Portland, OR: Author.
- McLaughlin, M.W. (1975). *Evaluation and reform*. Cambridge, MA: Ballinger Publishing Co.
- Weisberg, H.I. Short-term cognitive effects of Head Start programs: A report of the third year of planned variation—1971-72. Cambridge, MA: Huron Institute.

Project Follow Through: In-depth and Beyond

Gary Adams
Educational Achievement Systems, Seattle

The following article is a summary of a chapter in Adams, G., & Engelmann, S. (1996), Research on Direct Instruction.

Project Participants and Models

The Follow Through project was the largest, most expensive educational experiment ever conducted. This federal program was originally designed to be a service-oriented project similar to Head Start. However, because of funding cutbacks the emphasis was shifted from service to program evaluation. Over 75,000 low income children in 170 communities were involved in this massive project designed to evaluate different approaches to educating economically disadvantaged students from kindergarten through grade 3. State, school, and national officials nominated school districts that had high numbers of economically disadvantaged students. Parent representatives of these school districts chose to participate after hearing presentations from the 20 different program designers (sponsors). Each participating district implemented the selected sponsor's approach in one or more schools. For participating, each district received \$750 per student beyond the normal level of funding.

Each sponsor was required to:

- "provide the community with a well-defined, theoretically consistent and coherent approach that could be adapted to local conditions;
- provide the continuous technical assistance, training, and guidance necessary for local implementation of the approach;
- exercise a 'quality control' function by consistently monitoring the progress of program implementation;
- serve as an agent for change as well as a source of program consistency by asking the community in retaining a consistent focus on the objectives and requirements of

the approach rather than responding in an ad hoc manner to the daily pressures of project operations;

- ensure implementation of a total program, rather than a small fragment, such as reading, with a resulting possibility for a major impact on the child's life, and
- provide a foundation for comprehending and describing results of evaluation efforts" (Stebbins, St. Pierre & Proper, 1977, p. 5).

The orientation of the sponsors varied from the loosely-structured open classroom approach to the highly-structured behavior analysis approach. Nine of the original sponsors qualified for inclusion in the evaluation. To be included, a sponsor had to have more than three active sites that could be compared to control sites in the same communities.

Abt Associates used the system developed by White to classify the approaches of the different models. The first dimension was the theoretical orientation of the models:

- The behavioristic approach is based on the belief that all behaviors are learned. The reason that disadvantaged children are behind is because no one has taught them necessary social and academic skills. The training is based on selecting the behavioral objectives that are needed. Then teachers reinforce the steps in the behavioral objectives. The general label for this group became the *Basic Skills Models*.
- The cognitive development approach is based on the sequence of normal cognitive growth. The reason that disadvantaged children are behind is because they have

insufficient normal cognitive experiences. The orientation of this approach is to provide interactions between children and teachers. During these interactions, children learn how to solve problems and learn verbal skills based on a self-directed process. Emphasis is placed on the teacher providing age-appropriate cognitive materials and experiences. The general label for this group was the *Cognitive/Conceptual Skills Models*.

- The psychodynamic approach is based on the assumption that socioemotional development (the development of the "whole child") is essential to educational improvement. Emphasis is placed on trying to improve children's self-esteem and peer interactions. The goal for the teacher is to provide an environment in which children can move toward the goal of self-actualization through children making their own free choices. However, it is assumed that children know what is best for their personal growth. The general label for this group was the *Affective Skills Models*.

Basic Skills Models

Direct Instruction Model (University of Oregon)—Developed by Siegfried Engelmann and Wes Becker, this model used the DISTAR (DISTAR is an acronym for Direct Instruction System for Teaching And Remediation) reading, arithmetic, and language programs. The model assumes that the teacher is responsible for what the children learn.

Behavior Analysis Model (University of Kansas)—Developed by Donald Bushell, this model used a behavioral (reinforcement) approach for teaching reading, arithmetic, handwriting, and spelling. Social praise and tokens were given to the children for correct responses and the tokens were traded for desired activities. Teachers used programmed reading programs in which the task was presented in small steps. The instructional program was not specified by the model. Two sites used the DISTAR materials. Many used *Sullivan Programmed Phonics*. Students were monitored and corrective procedures were implemented to ensure student progress.

Language Development (Bilingual) Model (Southwest Educational Developmental Laboratory)—This curriculum-based model used an eclectic approach based on language development. When appropriate, material was presented first in Spanish and then in English.

Cognitive/Conceptual Skills Models

Cognitively-Oriented Curriculum (High Scope Foundation)—This popular program was directed by David Weikart and was based on Piaget's belief that there are underlying cognitive processes. Children were encouraged to schedule their own activities and then follow their schedules. The teacher modeled language through the use of labeling and explaining causal relationships. Also, the teacher fostered a positive self-concept through the way the students were given choices.

Florida Parent Education Model (University of Florida)—Based on the work of Ira Gordon, this program taught parents of disadvantaged children to teach their children. At the same time, students were taught in the classroom using a Piagetian approach. Parent trainers coordinated the teaching. Emphasis included not only language instruction, but also affective, motor, and cognitive skill instruction.

Tucson Early Education Model (University of Arizona)—Developed by Marie Hughes, TEEM used a language-experience approach (much like the whole language approach) that attempted to elaborate the child's present experience and interest. The model was based on the assumption that children have different learning styles so the child-directed choices are important. The teacher assists by helping children compare, recall, and locate relationships.

Affective Skills Models

Bank Street College Model (Bank Street College of Education)—This model used the traditional middle-class nursery school approach that was adopted by Head Start. Through the use of learning centers, children had many options, such as counting blocks and quiet areas of reading. The teacher is responsible for program implementation by taking advantage of learning situations. The classroom is structured to increase learning opportunities.

Open Education Model (Education Development Center)—Derived from the British Infant School model, this model focuses on building the children's responsibility for their own learning. Reading and writing were not taught directly, but through stimulating a desire to communicate.

Responsive Education Model (Far West Laboratory)—Developed by Glenn Nimick, this is an eclectic model using the work of O.K. Moore, Maria Montessori, and Martin Deutsch. The model used learning centers and the child's interests to determine when and where the child is stationed. The development of self-esteem is considered essential to the acquisition of academic skills.

Program Design

Each model had 4 to 8 sites with children that started school in kindergarten and some models also had sites with children that started in first grade. Each Follow Through (FT) school district identified a non-Follow Through (NFT) comparison school for each Follow Through site. The comparison school acted as a control group. Unfortunately, the NFT sites that were selected tended to have children who were less economically disadvantaged than the Follow Through sites. Because of this problem, Abt Associates used a covariance statistical analysis process to adjust for initial differences.

A total of 9,255 FT and 6,485 NFT children were in the final analysis group. Students in each school district site were tested at entry and then each spring until the third grade. The DI Model group included low income students in 20 communities. These communities varied widely—rural and urban—blacks, whites, Mexican Americans, Spanish American, Native Americans, and a diverse mixture of other ethnic groups.

The Stanford Research Institute was initially awarded a contract for data collection and Abt Associates received a contract for data analysis. The Office of Education determined the final design of the project with consultation from the Huron Institute. Because the sponsors had different approaches, the data collection was comprehensive. Assessment information was collected in the areas of basic skills (academic), cognitive, and affective behavior. The process of selecting appropriate assessment instruments was an arduous task given the time constraints of trying to select the most reliable, valid tests that could be administered in the least amount of time.

The following tests were used to assess basic skills, cognitive, and affective achievement: the Metropolitan Achievement Test (MAT), the Wide Range Achievement Test (WRAT), the Raven's Coloured Progressive Matrices, the Intellectual Achievement Responsibility Scale (IARS+ and IARS-), and the Coopersmith Self-Esteem Inventory. The MAT is a respected achievement test that assesses Basic Skills and Cognitive-Conceptual Skills. The Basic Skills scales of the MAT included Listening for Sound (sound-symbol relationships), Word Knowledge (vocabulary words), Word Analysis (word identification), Mathematic Computation (math calculations), Spelling, and Language (punctuation, capitalization, and word usage). The WRAT measured number recognition, spelling, word reading, and oral and written math problems.

The Cognitive Skills scales of the MAT included Reading (comprehension of written passages), Math-

ematics Concepts (knowledge of math principles and relationships), Mathematical Problem Solving (the use of reasoning with numbers). Also, the Raven's Coloured Progressive Matrices was used. The Raven's test, however, did not prove to discriminate between models or show change in scores over time.

Affective Skills was assessed using two instruments. The IARS was designed to assess whether children attribute their success (+) or failures (-) to themselves or external forces. The Coopersmith Self-Esteem Inventory is designed to assess how children feel about themselves, the way they think other people feel about them, and their feelings about school.

Comparisons Across Follow Through Sponsors

Students started in either kindergarten or first grade and were retested yearly through the end of third grade. While critics have complained about test selection and have usually suggested more testing, the assessment effort of this study was well beyond any other educational study conducted before, or since.

Significant Outcome Comparison

Abt Associates analyzed the data by comparing each Follow Through model's scores to both the local comparison group and the national pooled comparison group (created by combining the comparison groups from all nine Follow Through models). Local comparison scores and national pooled comparison scores were used as covariates to analyze each variable. A plus (+) was given if (a) the Follow Through (FT) group exceeded the Non-Follow Through (NFT) group by one-fourth standard deviation (.25 effect size) and (b) the difference was statistically significant. A minus (-) was given if the NFT score exceeded the FT score by one-fourth standard deviation (.25 effect size) and was statistically significant. If the results did not reach either the plus or the minus criterion, the difference was null and left blank.

The following index is based on a comparison of each model's site with the local and pooled national comparison groups. If either the pooled or local comparison were plus (+), the effect is recorded as a plus. If either or both was a minus (-), the effect is recorded as a minus. Then the plus and minus values are summed and multiplied by 100 so the possible range of scores was from -100 to 100. If the Follow Through model group scored consistently higher than the comparison group on a variable, then the index would be a positive number. If the

comparison group scored higher, the index would be negative. If there was no difference between the two groups, the score would be zero (0.00).

Figure 1 shows the results of this analysis. As you can see by the number of negative scores, the local or national pooled comparison group scores were *higher* than most Follow Through models.

Only the Direct Instruction model had positive scores on all three types of outcomes (Basic Skills, Cognitive, and Affective). Overall, the Direct Instruction model was highest on all three types of measures.

The results were very different from expectations suggested by the model orientations. The three programs in the Basic Skills model had the best basic skills, cognitive skills, and affective skills scores. Of the three orientations, the Basic Skills models (Direct Instruction, Behavior Analysis, and Southwest Lab) had the best basic skills scores. The Cognitive models (Parent Education, TEEM, and Cognitively-Oriented Curriculum) ranked second in cognitive skills scores; however, the average rank of 5.0 is far from the average rank of 2.8 for the Basic Skills model. The Affective Models had the worst affective ranks (6.7 compared to 2.7 for the Basic Skills models).

Figure 1 provides more details on the models' rankings. The DI model had, by far, the highest basic skills scores while the other two Basic Skills models had more modest results (the Behavior Analysis model had a slight positive score and the Southwest Labs model score was 0.0).

Figure 1 also shows that none of the Cognitive Models had positive cognitive scores. In fact, the

Direct Instruction Model was the only model of the nine that had a positive cognitive score (and the results were extremely positive - over 35%). In contrast, students in two of the three cognitively-oriented models [TEEM and Cognitive Curriculum (High Scope)] had the *lowest* cognitive scores.

Critics have often complained that the DI model was a pressure cooker environment that would negatively impact students' social growth and self-esteem. As the Abt Associates' authors note:

Critics of the model have predicted that the emphasis of the model on tightly controlled instruction might discourage children from freely expressing themselves and thus inhibit the development of self-esteem and other affective skills. (Stebbins, St. Pierre & Proper, p. 8)

Because of this expectation, the affective scores are of interest. Three of the five lowest scoring models on the affective domain were models that targeted improving affective behavior; none of the affective models had positive affective scores. In contrast, all Basic Skills models had positive affective scores with the Direct Instruction model achieving the highest scores. The theory that an emphasis on basic skills instruction would have a negative impact on affective behavior is not supported by the data. Instead, it appears that the models that focused on an affective education not only had a negative impact on their students' basic skills and cognitive skills, but also on their affective skills.

Fine Tuning the Results

The Bereiter-Kurland Reanalysis. A group funded by the Ford Foundation (House, Glass, McLean, & Walker, 1978) questioned certain aspects about the

0 = No difference between the Follow Through model and the comparison group

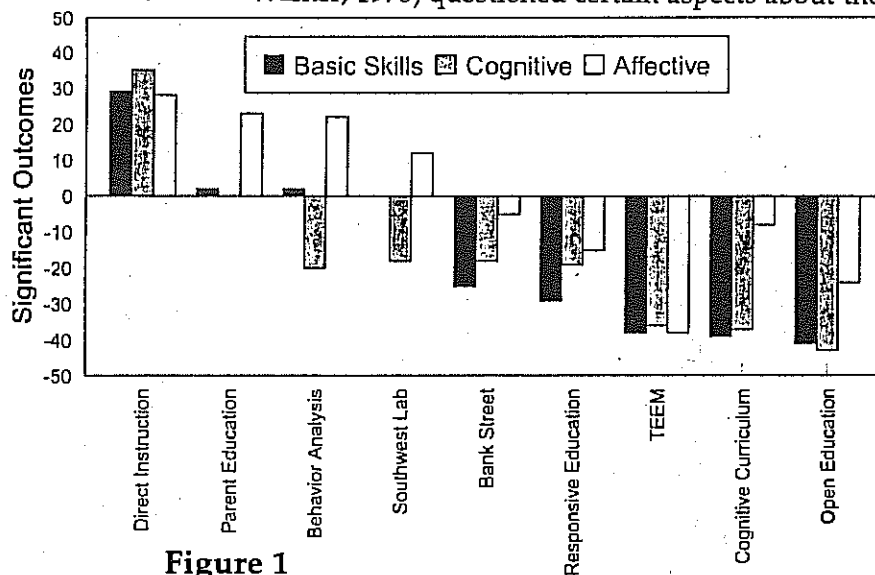


Figure 1

Significant outcomes comparison across Follow Through models.

(adapted from Stebbins et al., 1977)

test selection and data analysis in the Abt report. After reviewing the critiques of the Abt Report by House et al., 1978), Bereiter and Kurland (1981-1982) reanalyzed the data of that report based on the criticisms that the report used an inappropriate unit of measurement for the dependent variable and inappropriate covariates. The Bereiter-Kurland reanalysis was based on:

- Using the site means as the dependent variable.
- Using these site scores as covariates: socioeconomic status and ethnic and linguistic difference from the mainstream.
- Using only models that had data from 6 or more sites.

Each model had the possibility of 77 statistically significant differences (7 other models times 11 MAT subscale scores). Fifty of the 77 (65%) possible differences for the DI group were statistically significant based on Newman-Keuls Tests $p < .05$). In contrast, the Behavior Analysis group showed only 18 of 77 (23%) significant differences.

None of the other six models showed any statistically significant differences on any of the 11 MAT subscales (0 of 396 possible combinations). This means, for example, that none of the 11 MAT Bank Street scores differed significantly from any of the Responsive Education, TEEM, Cognitive Curriculum, Parent Education, or Open Education mean scores.

Another way of showing the difference between models was through the use of effect size comparisons. Figure 2 shows a different display of the information provided by Bereiter and Kurland (also Figure 2 in the Bereiter & Kurland review). In Figure 2, the effect size of the DI model is compared to the average effect size for the other Follow Through models. The differences are dramatic, even though the DI data include the Grand Rapids site that did not truly implement the DI model. The differences would be greater if only DI sites with implementation fidelity were included.

To provide a clearer picture of the differences, Figures 3-4 display the Bereiter-Kurland findings according to domain. First, Figure 3 shows a comparison of effects for the Basic Skills scores between the DI group and the average effect size of the other Follow Through groups. Remember an effect size of .25 is thought to be educationally significant. Differences in some MAT Basic Skills subscales scores are over 3.0 (Total Language and Language B). The

average difference in Basic Skills scores between Direct Instruction and the other models was 1.8.

Figure 4 shows the differences in the cognitive scores between the DI models and the average Follow Through model. Effect sizes are above 1.0 for all but one difference.

Overall, the Bereiter-Kurland reanalysis provides even stronger support for the effectiveness of Direct Instruction. As the authors noted, only the DI and Behavior Analysis models had positive results and the DI model results were vastly superior.

Changing the Abt Report Criteria

Becker and Carnine (1981) had two other complaints about the Abt Associates report, which resulted in the report *underrepresenting* the superiority of the DI model. First, because of the problem of mismatches between comparison groups that initially had higher entry scores than the Follow Through model groups, Abt Associates deleted these data from subsequent analyses. Unfortunately for the DI model, sometimes the scores for the comparison groups were significantly higher at entry, but by the end of third grade the DI group scored significantly higher than the comparison groups. Abt Associates decided to delete these groups because of the initial entry differences. Also, data were excluded if there were significant differences between the two groups in preschool experience per site, even though preschool experience (e.g., Head Start) had only a very low correlation with later achievement (-0.09). (This variable was not used in the previously cited Bereiter-Kurland study.) Overall, approximately one-third of the data was excluded from most Follow Through models because of these decision rules.

Figures 5-7 show the differences in results based on these analyses. When data were kept for sites where there were initial performance differences, the highest scoring model (DI) scored even higher whereas the lower scoring models (Cognitive Curriculum and Open Education) scored even lower. The scores for the other models stayed roughly the same.

Becker and Carnine re-analyzed the data without the Grand Rapids site. The Grand Rapids site stopped using Direct Instruction when there was a change in program director. Even though this problem was well documented, Abt Associates included the Grand Rapids site in the DI data. Becker and Carnine reanalyzed the Abt Associates results without the Grand Rapids site. Figures 5-7 shows the already high scores for the DI group became even higher when the Grand Rapids data were removed.

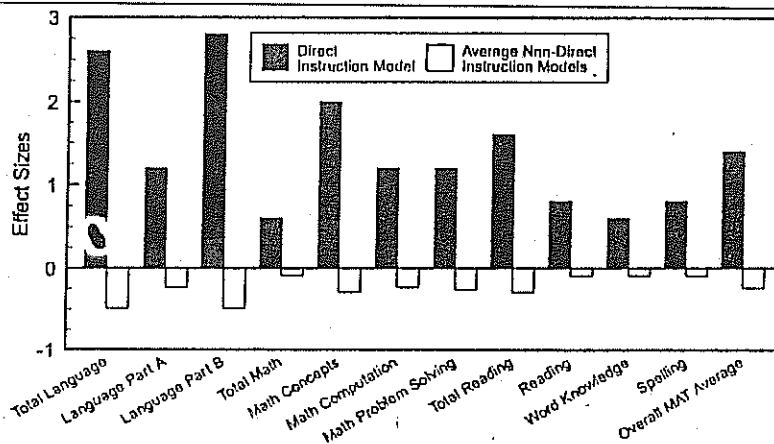


Figure 2

This comparison includes the Grand Rapids site.
(Adapted from Bereiter & Kurland, 1981-1982)

Effect size comparison (DI to other models).

Norm-Referenced Comparisons

Another way of looking at the Abt Associates data is to compare median grade-equivalent scores on the normed-referenced Metropolitan Achievement Test that was used to evaluate academic progress. Unlike the previous analysis that compared model data to local and pooled national sites, the following norm-referenced comparisons show each model's MAT scores based on the MAT norms. Figure 8 shows the results across four academic subjects. The comparisons are made to a baseline rate of the 20th percentile which was the average expectation of disadvantaged children without special help. The figure displays the results in one-fourth standard deviation intervals.

Clearly, children in the DI model showed consistently higher scores than the other models, and also

the students in the Southwest Lab and the Open Education model were below expected levels of achievement based on norms of performance in traditional schools in all four academic subjects.

Only three of 32 possible reading scores of the other eight models were above the 30th percentile. The DI students scored 7 percentile points higher than the second place group (Behavior Analysis) and over 20 percentile points higher than the Cognitive Curriculum (High Scope), Open Education, and Southwest Lab Models.

Except for children in the DI model, the math results are consistently dismal. The only other model above the 20th percentile was the Behavior Analysis model. DI students scored 20 percentiles ahead of the second place group (Behavior Analysis) and 37 percentiles higher than the last place group (Cognitive Curriculum/High Scope).

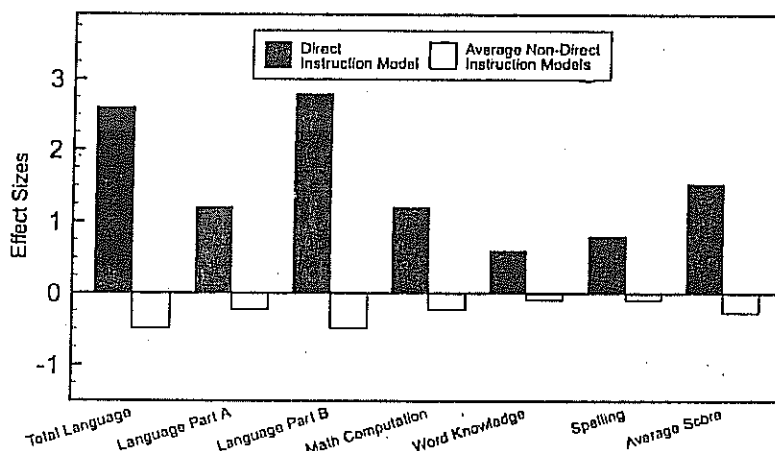


Figure 3

This comparison includes the Grand Rapids site.
(Adapted from Bereiter & Kurland, 1981-1982)

Bereiter analysis of basic skills Abt data.

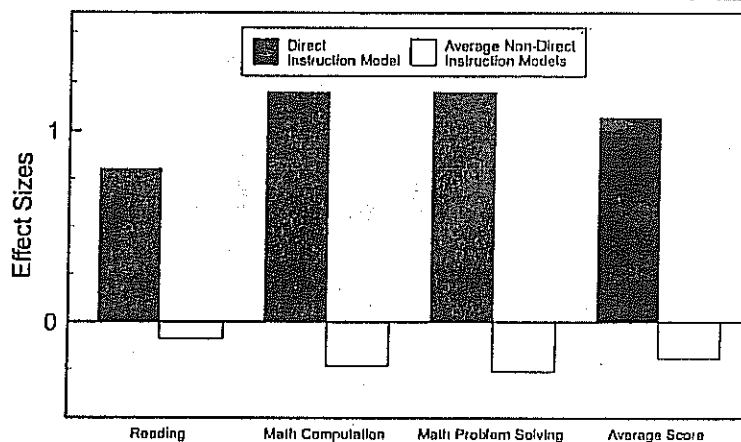


Figure 4

Bereiter analysis of cognitive ability Abt data.

This comparison includes the Grand Rapids site.
(Adapted from Bereiter & Kurland, 1981-1982)

In spelling, the DI model and the Behavior Analysis model were within the normal range. DI students scored 2 percentiles above the second place group (Behavior Analysis), 19 percentiles above the third place group, and 33 percentiles above the last place group (Open Education).

Like the previous academic subjects, the DI model was clearly superior in language. DI students scored 29 percentiles above the second place group (Behavior Analysis) and 38 percentiles above the last place group (Cognitive Curriculum/High Scope).

Conclusions

For many people the use of normed scores are more familiar than the use of the index described in the previous section. No matter which analysis is used, children who were in the DI model made the

most gains when compared to the other eight models. With the possible exception of the Behavior Analysis model, all other models seem to have little positive effect on the academic progress of their children.

The increase amounts of money, people, materials, health and dental care, and hot lunches did not cause gains in achievement. Becker (1978) observed that most Follow Through classrooms had two aides and an additional \$350 per student, but most models did not show significant achievement gains.

Popular educational theories of Piaget and others suggest that children should interact with their environment in a self-directed manner. The teacher is supposed to be a facilitator and to provide a responsive environment. In contrast, the successful DI model used thoroughly field-tested curricula that

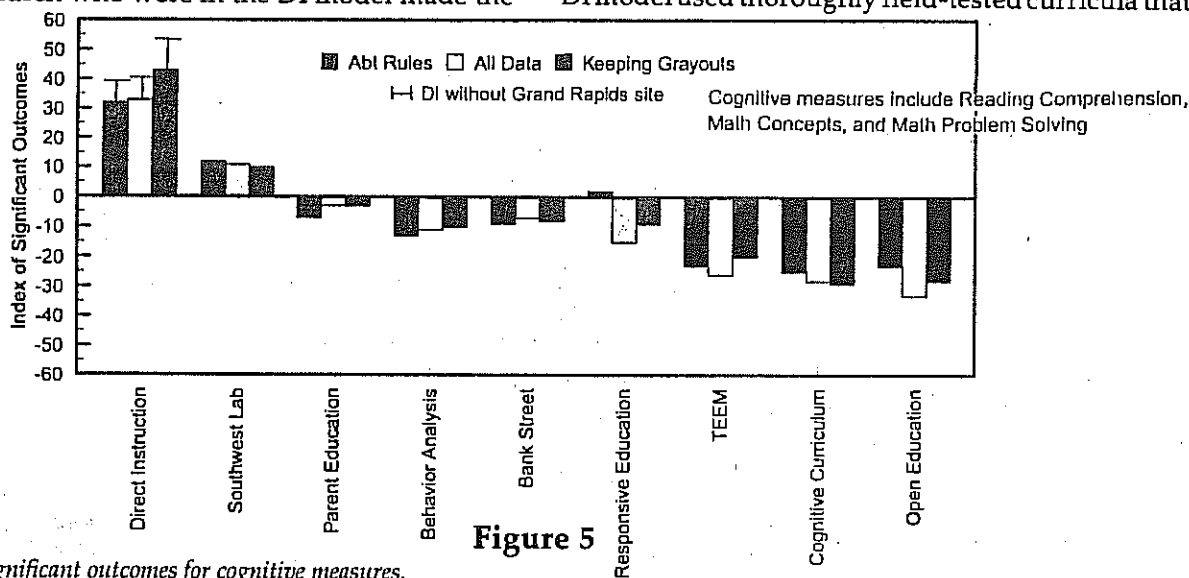


Figure 5

Index for significant outcomes for cognitive measures.

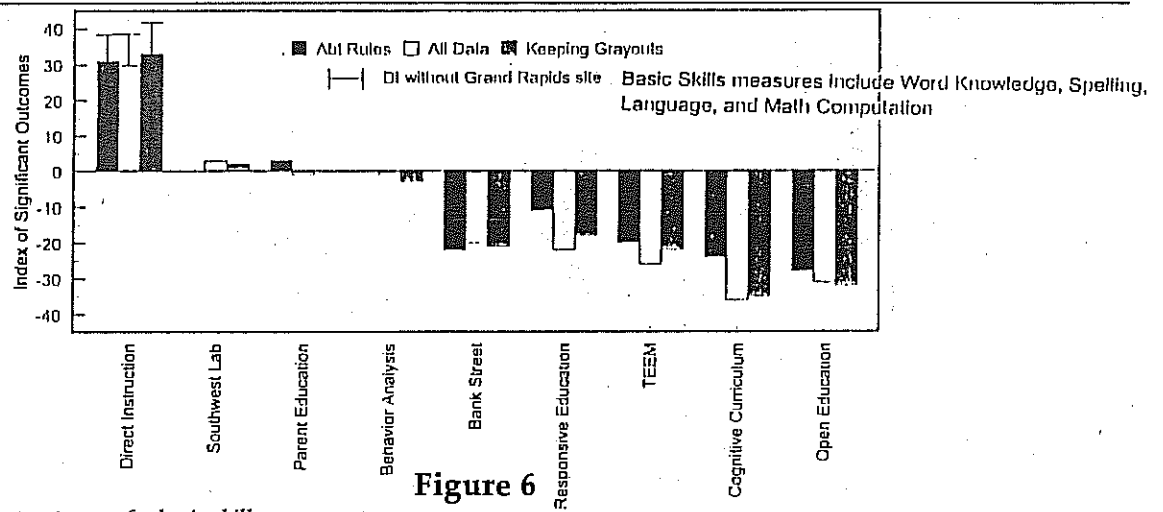


Figure 6

Index for significant outcomes for basic skills measures.

teachers should follow for maximum success. The Follow Through models that were based on a self-directed learner model approach were at the bottom of academic and affective achievement. The cognitively-oriented approaches produced students who were relatively poor in higher-order thinking skills and models that emphasized improving students' self-esteem produced students with the poorest self-esteem.

Subsequent Analyses

Variability Across DI Sites

The Abt Associates findings were criticized by House, Glass, McLean, & Walker (1978) and then defended by others (Anderson, St. Pierre, Proper, Stebbins, 1978; Becker, 1977; Bereiter & Kurland,

1981-82; Wisler, Burns, & Iwanoto, 1978). One Abt Associates finding was that there was more variability within a model than between models.

This statement is consistent with the often cited belief that "Different programs work for different children" or another way of saying "Not all programs work with all children." The following sections provide research results that contradict this statement. The problem is that the statement doesn't match the data.

Gersten (1984) provided an interesting picture of the consistency of achievement scores of urban DI sites after the Abt report was completed. Figure 9 shows the results in 3rd grade reading scores from 1973 to 1981 in four urban cities. The reading scores are consistently around the 40th percentile. Based on non-Follow Through students in large North-

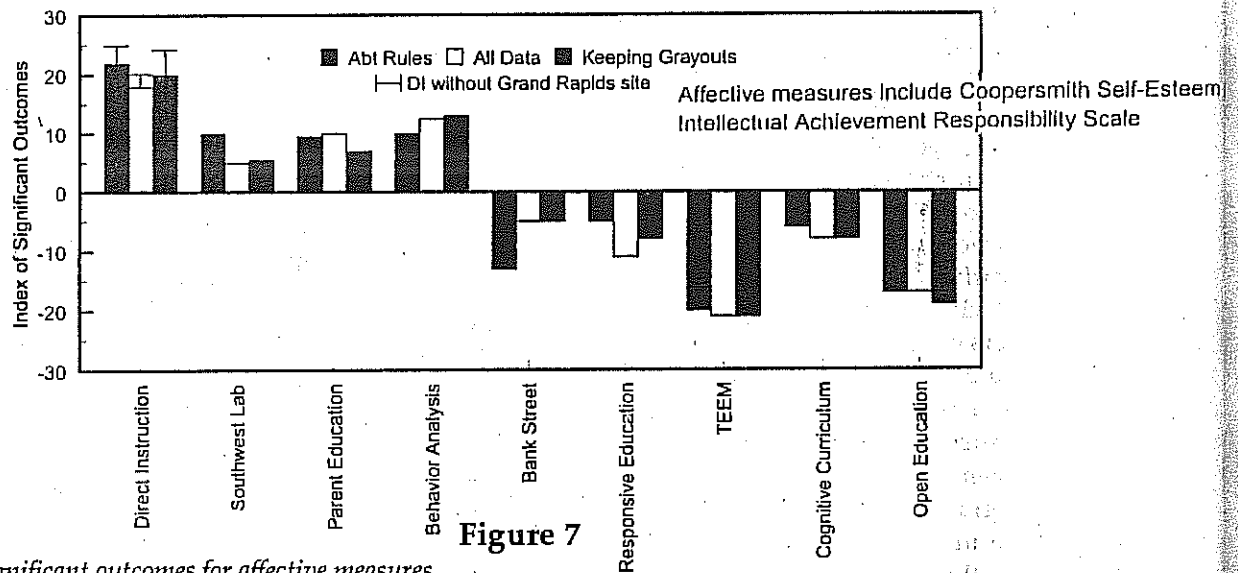


Figure 7

Index for significant outcomes for affective measures.

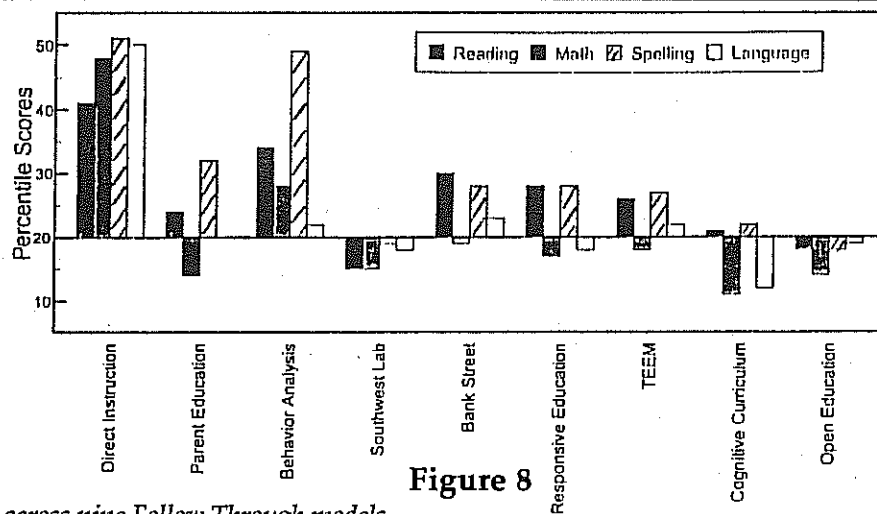


Figure 8

Percentile scores across nine Follow Through models.

west urban settings, the expected score is the 28th percentile on the MAT. Some variability is due to the differences between tests when some districts changed tests over the nine year period. Also, Gersten mentioned that the drop in the New York scores in 1978 and 1979 may have been because of budgetary reductions during those years.

Figure 10 shows the stability of math scores. The math scores for these three sites tend to be consistently in the 50th percentile range. New York did not collect information of math during this period. Based on the math scores of large Northwest cities, non-Follow Through students would be expected to score at the 18th percentile.

Follow-Up Studies

Fifth and Sixth Grade Follow-up

Some critics of DI have indicated that many, if not most, early DI achievement gains will disappear over time. There are different reasons given for this prediction. One reason given is that the DI students were "babied" through sequences that made instruction easy for them. They received reinforcement and enjoyed small group instruction, but they would find it difficult to transition to the realities of the "standard" classroom.

DI supporters give different reasons for suggesting that DI results would decrease over time. The DI students were accelerated because they had been taught more during the available time than they would have been taught during the same time in a traditional program. Upon leaving Follow Through, they would be in instructional settings that teach relatively less than the Follow Through setting achieved. Statistically, there would be a tendency for them to have a regression toward the mean

effect. Phenomonologically, students would be provided with relatively fewer learning opportunities and would tend to learn less accordingly.

In any case, the effects observed at a later time are not the effects of Follow Through. They are the effects of either three or four years of Follow Through and the effects of intervening instructional practices. Engelmann (1996) observed that because the typical instruction provided for poor children in grades 4 and beyond has not produced exemplary results, there is no compelling reason to use results of a follow-up to evaluate anything but the intervening variables and how relatively effective they were in maintaining earlier gains.

Junior and Senior High School Follow-up

New York City Follow-up

One of the most interesting long-term follow-up studies was conducted by Linda Meyer (1984). She tracked students from two schools in Ocean Hill-Brownsville section of Brooklyn. This district was one of the lowest of the 32 New York school districts. The fifteen elementary schools in District 23 had an average rank 519th out of the 630 elementary schools.

PS 137 was the only DI Follow Through site in New York City. Meyer selected a comparison school that matched the DI school on many variables. Over 90% of the students were minority students and over 75% were from low-income families.

Meyer retrieved the rosters of the first three cohort groups (1969, 1970, and 1971) and included students who received either three or four years DI instruction. With the cooperation of the New York City Central Board of Education and the Office of the Deputy Chancellor for Instruction, students were located through the computer database. Meyer and staff were able to locate 82% of the former DI stu-

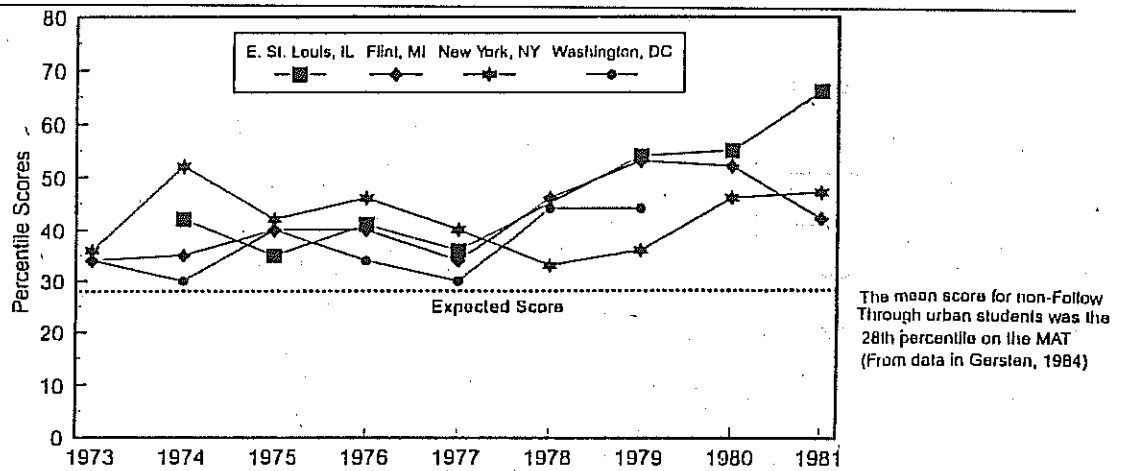


Figure 9

Total reading scores for K-3 students. Stability of effects: Percentile equivalents at the end of Grade 3.

dents and 76% of the control students. These rates should be considered high because it would be expected that over time many students would move totally out of the area.

Table 1 shows the grade equivalent scores for the DI and comparison groups of the three cohort groups. At the end of 9th grade, the three DI groups were on average one year above the three comparison groups in reading (9.20 versus 8.21) ($p < .01$) with an effect size of .43. In math, the DI groups were approximately 7 months ahead of the comparison group (8.59 versus 7.95) which was not statistically significant ($p < .09$), but educationally significant based on an effect size of .28.

Achievement Growth for Other Sites

Gersten, Keating, and Becker (1988) provide similar information for other sites. Table 2 shows the effect sizes of the E. St. Louis and Flint sites at the end of ninth grade. Most effect sizes were above the .25 level of being educationally significant. It should be noted that the 3-K East St. Louis group that started in kindergarten, instead of first grade, had four years of instruction (not three) had the second highest effect size (.49).

Table 3 shows similar effectiveness in the math. The results of these two analyses clearly show that while the superiority of DI diminishes with the time spent in traditional curricula, the advantage of the

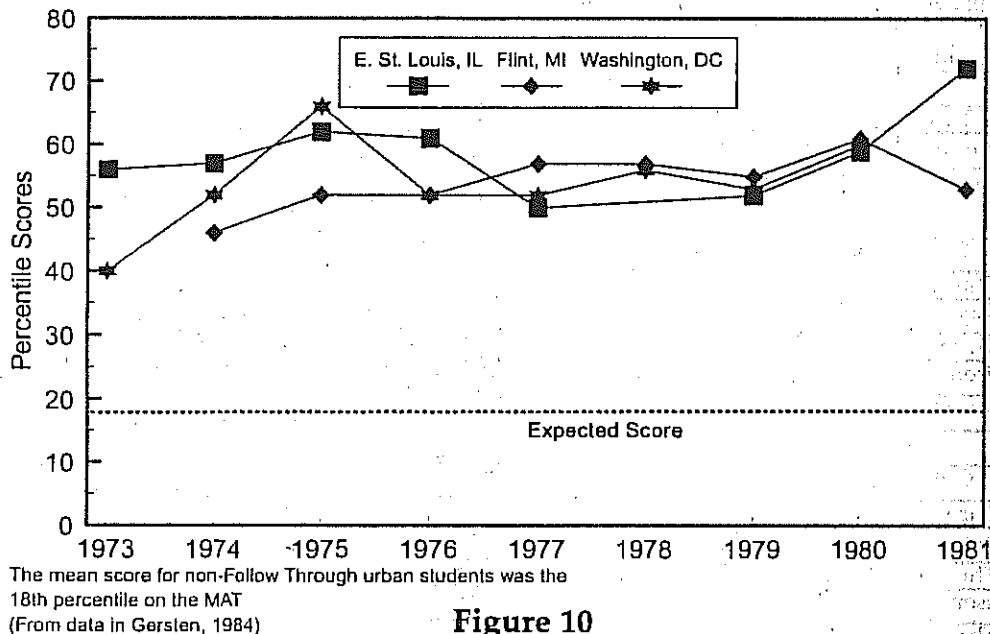


Figure 10

Total math scores for K-3 students. Stability of effects: Percentile equivalents at the end of Grade 3.

Table 1
Results of T-Tests Comparisons and Effect Sizes for Reading
and Math at the End of 9th Grade

| Subject and Cohort | | Direct Instruction | Comparison Group | | |
|--------------------|----------------------|--------------------|------------------|------|-------------|
| | | Grade Equivalent | Grade Equivalent | p | Effect Size |
| Reading | Combined 1, 2, and 3 | 9.2 (N=62) | 8.21 (N=93) | <.01 | .43 |
| Math | Combined 1, 2, and 3 | 8.59 (N=63) | 7.95 (N=93) | <.09 | .28 |

DI lasts. Educational significant differences occur in reading (overall effect size = .43) and in math (overall effect size = .25).

Graduation Rates and College Acceptance Rates at Other Sites

Darch, Gersten, & Taylor (1987) tracked Williamsburg (SC) students in the original Abt study (students entering first grade in 1969 and 1970) to compare graduation rates. All students were black

and had stayed in the county school system. Table 4 shows a statistically significant difference in drop-out rate for Group 1 (the 1969 group), but the difference in drop-out rate was not statistically significant for Group 2 (the 1970 group).

A total of 65.8% of the Group 1 Follow Through students graduated on time in contrast to 44.8% of the comparison group (a statistically significant difference - $p < .001$). For Group 2, 87.1% of the Follow Through group and 74.6% of the comparison group graduated on time (a nonsignificant statistical dif-

Table 2
Ninth Grade Reading Achievement Results from East St. Louis, Flint, and New York

| Site (Test Used) | Sample Size | %tile | Grade Equival. | p | Effect Size |
|--|-------------|-------|----------------|-----|-------------|
| E. St. Louis (California Achievement Test) | | | | | |
| 1-Follow Through | 118 | 22nd | 7.2 | .05 | .20 |
| Comparison | 158 | 17th | 6.7 | | |
| 2-Follow Through | 64 | 20th | 7.0 | .05 | .25 |
| Comparison | 168 | 14th | 6.3 | | |
| 3-Follow Through | 59 | 23rd | 7.4 | .05 | .31 |
| Comparison | 138 | 18th | 6.9 | | |
| 3-K-Follow Through | 54 | 40th | 8.8 | .01 | .49 |
| Comparison | 121 | 26th | 7.6 | | |
| Average across cohorts | | | | | .33 |
| Flint (SRA Test) | | | | | |
| 1-Follow Through | 74 | 29th | X | NS | .25 |
| Comparison | 27 | 22nd | X | | |
| 3-Follow Through | 51 | 34th | X | .15 | .28 |
| Comparison | 28 | 25th | X | | |
| Average across cohorts | | | | | .27 |
| New York City (California Achievement Test) | | | | | |
| 1-Follow Through | 18 | 50th | 9.6 | .01 | .83 |
| Comparison | 26 | 27th | 7.6 | | |
| 2-Follow Through | 17 | 39th | 8.5 | NS | .27 |
| Comparison | 30 | 31st | 8.0 | | |
| 3-Follow Through | 29 | 48th | 9.4 | NS | .19 |
| Comparison | 37 | 43rd | 8.9 | | |
| Average across cohorts | | | | | .43 |

Table 3
Ninth Grade Math Achievement Results from East St. Louis, Flint, and New York

| Site (Test Used) | Sample Size | %tile | Grade Equiv. | <i>p</i> | Effect Size |
|--|-------------|-------|--------------|----------|-------------|
| E. St. Louis (California Achievement Test) | | | | | |
| 1-Follow Through | 114 | 18th | 7.3 | .05 | .26 |
| Comparison | 155 | 14th | 6.8 | | |
| 2-Follow Through | 58 | 15th | 6.9 | NS | .09 |
| Comparison | 154 | 13th | 5.9 | | |
| 3-Follow Through | 58 | 19th | 7.4 | NS | .12 |
| Comparison | 154 | 18th | 7.2 | | |
| 3-K-Follow Through | 53 | 30th | 8.3 | .01 | .46 |
| Comparison | 117 | 20th | 7.5 | | |
| Average across cohorts | | | | | .20 |
| Flint (SRA Test) | | | | | |
| 1-Follow Through | 72 | 30th | X | NS | -.17 |
| Comparison | 28 | 35th | X | | |
| 3-Follow Through | 51 | 50th | X | .05 | .59 |
| Comparison | 28 | 31st | X | | |
| Average across cohorts | | | | | .21 |
| New York City (California Achievement Test) | | | | | |
| 1-Follow Through | 17 | 36th | 8.3 | NS | .21 |
| Comparison | 27 | 31st | 8.0 | | |
| 2-Follow Through | 17 | 34th | 8.1 | NS | .23 |
| Comparison | 29 | 27th | 7.6 | | |
| 3-Follow Through | 29 | 44th | 9.0 | NS | .31 |
| Comparison | 38 | 36th | 8.3 | | |
| Average across cohorts | | | | | .25 |

ference). Also, 27% of the Group 1 Follow Through were accepted into college in contrast to 13% of the comparison group; the difference for Group 2 in college admission was not significant.

Meyer, Gersten, & Gutkin (1983) calculated the rates of graduation, retention, dropping out, applying to college, and acceptance to college for the three cohort groups in the New York City site. Statistical analyses showed that the DI group had statistically significantly higher rates of graduation ($p < .001$), applying to college ($p < .001$), acceptance to college ($p < .001$) and lower rates of retention ($p < .001$) and dropping out ($p < .001$). The differences in graduation rates were consistent across the three cohort groups with over 60% of the DI students graduating in contrast to less than a 40% graduate rate for the three comparison groups. Meyer mentioned in her report that the difference in retention rate between Cohort II and Cohorts I and III may have been due to the principal retaining all students below grade level one year.

Conclusions

Educational reformers search for programs that produce superior outcomes with at-risk children, that are replicable and can therefore be implemented reliably in given settings, and that can be used as a basis for a whole school implementation that involves all students in a single program sequence, and that result in students feeling good about themselves. The Follow Through data confirm that DI has these features. The program works across various sites and types of children (urban blacks, rural populations, and non-English speaking students). It produces positive achievement benefits in all subject area — reading, language, math, and spelling. It produces superior results for basic skills and for higher-order cognitive skills in reading and math. It produces the strongest positive self-esteem of the Follow Through programs.

Possibly, the single feature that is not considered by these various achievements is the implied level of efficiency of the system. Some Follow Through

sponsors performed poorly in math, because they spent very little time on math. Most of the day focused on reading and related language arts. Although time estimates are not available for the various sponsors, some of them spent possibly twice as much time on reading as DI sites did. Even with this additional time, these sites achieved less than the DI sites achieved. For a system to achieve first place in virtually every measured outcome, the system is required to be very efficient and use the limited amount of student-contact time to produce a higher rate of learning than other approaches achieve. If the total amount of "learning" induced over a four-year period could be represented for various sponsors, it would show that the amount of learning achieved per unit of time is probably twice as high for the DI sites as it is for the non-DI sponsors.

Perhaps the most disturbing aspect of the Follow Through results is the persistence of models that are based on what data confirms is whimsical theory. The teaching of reading used by the Tucson Early Education Model was language experience, which is quite similar in structure and procedures to the whole language approach. The fact that TEEM performed so poorly on the various measures should have carried some implications for later reforms; however, it didn't. The notion of the teacher being a facilitator and providing children with incidental teaching was used by the British infant school model (Open Education). It was a flagrant failure, an outcome that should have carried some weight for the design of later reforms in the US. It didn't. Ironically, it was based on a system that was denounced in England by its Department of Science and Education in 1992. At the same time, states like California, Pennsylvania, Kentucky, Ohio, and others were in full swing in the National Association for the Education of Young Children's idiom of "developmentally appropriate practices," which are based on the British system.

Equally disturbing is the fact that while states like California were immersed in whole language and developmentally appropriate practices from the 1980s through mid 1990s, there was no serious attempt to find models or practices that work. Quite the contrary, DI was abhorred in California and only a few DI sites survived. Most of them did through deceit, pretending to do whole language. At the same time, those places that were implementing the whole language reading and the current idiom of math were producing failures at a tragic rate.

Possibly the major message of Follow Through is that there seems to be no magic in education. Gains are achieved only by starting at the skill level of the children and carefully building foundations that support higher-order structures. Direct Instruction has no peer in this enterprise.

References

- Adams, G., & Engelmann, S. (in press) *Research on Direct Instruction*. Seattle, WA: Educational Achievement Systems.
- Anderson, R., St. Pierre, R., Proper, E., & Stebbins, L. (1978). Pardon us, but what was the question again? A response to the critique of the Follow Through evaluation. *Harvard Educational Review*, 48(2), 1621-170.
- Becker, W. (1977). Teaching reading and language to the disadvantaged—what we have learned from field research. *Harvard Education Review*, 47, 518-543.
- Becker, W. C. (1978). National Evaluation of Follow Through: Behavior-theory-based programs come out on top. *Education and Urban Society*, 10, 431-458.
- Becker, W., & Carnine, D. (1981). Direct Instruction: A behavior theory model for comprehensive educational intervention with the disadvantaged. In S. Bijon (Ed.) *Contributions of behavior modification in education* (pp. 1-106), Hillsdale, NJ: Laurence Erlbaum.

Table 4
Longitudinal Follow-up Study: Percentage of Graduates and Dropouts for Direct Instruction Follow Through and Local Comparison Groups

| Group | Sample Size | % Graduated | % Drop-Out | % Still in School | p |
|---------------------|-------------|-------------|------------|-------------------|-----|
| 1-DI Follow Through | 131 | 93.1 | 6.1 | 0.8 | .01 |
| Comparison Group | 71 | 81.7 | 15.5 | 2.8 | |
| 2-DI Follow Through | 99 | 85.9 | 8.1 | 6.1 | NS |
| Comparison Group | 71 | 83.1 | 7.0 | 9.9 | |

Table 5
Percentages of Cohorts 1, 2, and 3 Students: Graduated High School, Retained, Dropped Out, Applied to College, and Accepted to College.

| | Graduated | Retained | Dropped Out | Applied to College | Accepted to College |
|-------------------------------|-----------|----------|-------------|--------------------|---------------------|
| Cohorts I, II, and III | | | | | |
| Direct Instruction | 62 | 21 | 28 | 34 | 34 |
| Comparison Group | 38 | 22 | 46 | 19 | 17 |
| Cohort I | | | | | |
| Direct Instruction | 62 | 4 | 39 | 50 | 41 |
| Comparison Group | 38 | 42 | 62 | 18 | 9 |
| Cohort II | | | | | |
| Direct Instruction | 64 | 50 | 20 | 23 | 23 |
| Controls | 38 | 35 | 58 | 25 | 21 |
| Cohort III | | | | | |
| Direct Instruction | 62 | 10 | 24 | 38 | 38 |
| Comparison Group | 39 | 21 | 18 | 24 | 21 |

Bereiter, c., & Kurland, M. (1981-82). A constructive look at Follow Through results. *Interchange*, 12, 1-22.

Darch, C., Gersten, R., & Taylor, R. (1987). Evaluation of Williamsburg County Direct Instruction Program: Factors leading to success in rural elementary programs. *Research in Rural Education*, 4, 111-118.

Gersten, R. (1984). Follow Through revisited: Reflections on the site variability issue. *Educational Evaluation and Policy Analysis*, 6, 411-423.

Gersten, R., Keating, T., & Becker, W. (1988). The continued impact of the Direct Instruction model: Longitudinal studies of Follow Through students. *Education and Treatment of Children*, 11(4), 318-327.

House, E., Glass, G., McLean, L., & Walker, D. (1978). No simple answer: Critique of the FT evaluation. *Harvard Educational Review*, 48(2), 128-160).

Meyer, Gersten, & Gutkin, (1983). Direct Instruction: A Project Follow Through success story in an inner-city school. *Elementary School Journal*, 84, 241-252.

Meyer, L. A. (1984). Long-term academic effects of the Direct Instruction Project Follow Through, *Elementary School Journal*, 84, 380-394.

Stebbins, L. B., St. Pierre, R. G., & Proper, E. C. (1977). Education as experimentation: A planned variation model (Volume IV-A & B) Effects of follow through models. Cambridge, MA.: Abt Associates.

Wisler, C., Burns, G.P., Jr., & Iwamoto, D. (1978). FT redux: A response to the critique by House, Glass, McLean, & Walker. *Harvard Educational Review*, 48(2), 171-185).

For information on ordering *Research on Direct Instruction*, contact Educational Achievement Systems, 319 Nickerson St., Suite 112, Seattle, WA 98109. Phone or Fax (206) 820-6111.

Follow Through: Why Didn't We?

Cathy L. Watkins
California State University, Stanislaus

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 approaches for educating children not been widely disseminated? Why has the knowledge gained from the Follow Through evaluation not been used to reform education in America? The answers to these questions may be found in part by looking at how instruction is viewed by the various elements of the educational establishment.

Follow Through provides an opportunity for such an analysis because it revealed how the educational industry collectively conceived of, planned, conducted, and interpreted a large scale educational experiment. When I was in graduate school, I wrote a lengthy paper in which I traced the history of Project Follow Through, looked at its implications for education, and analyzed the contingencies that determine educational practices. This article is condensed from the paper, which will be published this fall by the Cambridge Center for Behavioral Studies. I made a vow that I would tell the story of Project Follow Through to anyone who would listen.

History

Many people know the history of Project Follow Through far better than I, because they lived it. As I understand it, it goes something like this. In 1964 Congress passed the Economic Opportunity Act, which initiated a range of programs intended to fight poverty. The federal policy that emerged from the EOA was influenced by a growing consensus that education would be the antidote to poverty by providing skills necessary to break out of the exiting cycle. One of the best known programs to develop from this rationale was Head Start.

Head Start began in the summer of 1965. It was immediately popular and continues to enjoy tremendous public support today. It is commonly believed that Follow Through received impetus from the success of Head Start and from a study showing that gains made by Head Start children dissipated when they began school. In reality, the decision to initiate Follow Through was probably a function of both conviction and expediency. In any event, in February 1967, President Johnson requested that

Congress establish a program to "follow through" on Head Start. The outcome was Public Law 90-92, authorizing Project Follow Through. Although it has been referred to as the largest and most expensive educational experiment, Follow Through was not initially conceived as an experiment, but as a comprehensive social services program. However, before the program got underway, budget cuts forced a reconceptualization and Follow Through was converted to a longitudinal experiment aimed at finding effective methods for teaching disadvantaged children. The Follow Through experiment involved close to 10,000 children from 120 communities each year from 1968 to 1976. Follow Through continued as a service program until funding was eliminated in 1995.

Why has the knowledge gained from the Follow Through evaluation not been used to reform education in America? The answers to these questions may be found in part by looking at how instruction is viewed by the various elements of the educational establishment.

Design

The design of the Follow Through experiment was called planned variation. Based on the notion that a variety of curricula and instructional methods could be designed, implemented and evaluated, the planned variation approach was intended to reveal differences in effectiveness among different teaching approaches.

A plan was devised that made it possible to implement a variety of educational models in local school districts, while avoiding the appearance of unwarranted federal intervention. The Office of Education contracted with developers of educational approaches who then acted as sponsors of their model and worked cooperatively with districts to implement the model in Follow Through classrooms.

Although it has been referred to as the largest and most expensive educational experiment, Follow Through was not initially conceived as an experiment, but as a comprehensive social services program.

Each sponsor was responsible for translating the model's particular approach to education into practice. This included selecting or developing instructional materials, and training teachers in the model's instructional method. The Follow Through sponsors' task of designing a complete curriculum for the entire school day, had never before been attempted in educational reform.

The selection of sponsors began in January of 1968. Individuals or groups who were involved in developing new approaches for teaching young children were invited to present information about their programs. Sixteen developers subsequently submitted formal proposals, twelve of which were accepted. The approaches represented the entire spectrum of assumptions about instruction, ranging from the carefully controlled approach of the Direct In-

The design of the Follow Through experiment was called planned variation. Based on the notion that a variety of curricula and instructional methods could be designed, implemented and evaluated, the planned variation approach was intended to reveal differences in effectiveness among different teaching approaches.

struction and Behavior Analysis models to child-centered approaches such as Bank Street and Open Education. Ten additional sponsors were added over the following three-year period, not because they offered unique approaches to compensatory education, but because they offered the possibility of enlarging the Follow Through constituency.

The selection of sites progressed synchronously with sponsor selection. From among a group of 225 nominated school districts, a total 51 were selected, based on their ability to begin a comprehensive

services program before the start of the school year, their willingness to participate in the planned variation experiment and their working relationship with local community action agencies.

Sites and sponsored models were paired during a four day conference held in Kansas City in February, 1968. In an effort to increase cooperation in implementing the various models, local representatives were encouraged to choose the model they believed was most compatible with the goals and interests of their district. Each model was implemented in a variety of sites, where children received daily instruction in the model. Performance data were collected when children entered the program and at the end of each school year until they completed third grade.

Evaluation

The evaluation of this enormous project was complex and expensive. The data were collected by Stanford Research Institute and analyzed by Abt Associates.

Eleven outcome measures were included in the national evaluation. All sponsors agreed upon the outcome measures, which were intended to assess performance in different learning domains including basic academic skills, general problem-solving skills, and the development of self-concept.

For evaluation purposes Abt Associates divided models into three broad categories according to their areas of primary emphasis. The typology was determined based on the sponsor's own program description and stated goals and objectives of the models. The Basic Skills category included models that focused primarily on directly teaching fundamental skills in reading, arithmetic, spelling and language. The Cognitive-Conceptual category included that intended 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, "learning-to-learn" skills. Nine of the major models included in the national evaluation are described by model type in Table 1.

For each outcome subtest, Abt evaluators compared the performance of a group of Follow Through children at a given site with a comparison group. This process resulted in more than 2,000 comparisons. The difference between a Follow Through group and the comparison group was used as the measure of effect. An effect was judged to be educationally meaningful if the difference 1) was statistically significant and 2) was at least one quarter standard deviation. When Follow Through scores

exceeded non-Follow Through scores the outcome was considered positive. When non-Follow Through scores surpassed Follow Through scores, the outcome was considered negative. Average effects were computed for individual models, as well as for model types.

Eleven outcome measures were included in the national evaluation. All sponsors agreed upon the outcome measures, which were intended to assess performance in different learning domains including basic academic skills, general problem-solving skills, and the development of self-concept.

Abt Associates produced yearly reports, which were published in four volumes titled *Education as Experimentation: A Planned Variation Model*. Volume IV (Stebbins, et. al., 1977), provides the most comprehensive evaluation of the differential effectiveness of the models. The following findings of average effects for model types are paraphrased (in italics) from Volume IV-A (pp. 135-148).

Models that emphasized basic skills succeeded better than other models in helping children gain these skills. 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 was concerned with teaching skills such as spelling, math computation, language, and word knowledge. Note that the Abt report refers to the superiority of a model type. However, it is not inclusion in a category that leads to educational effectiveness, but the particular instructional materials and procedures used. The Direct Instruction model had an unequivocally higher average effect on scores in the basic skills domain than did any other model.

Where models emphasized other skills, the children they served tended to score lower on tests of basic skills than they would have done without Follow Through. With the exception of the Florida Parent Education model, all Cognitive-Conceptual and Affective-Cognitive models had more negative than positive outcomes on measures in the basic skills domain. That is, performance of students in the comparison group was superior to that of the Follow Through students in those models. At the end of third grade, children taught in these models had achievement scores that

were, in fact, lower than would have been predicted in the absence of "compensatory" education. Thus, four years spent in the majority of models actually increased the educational deficits that Follow Through was intended to remediate.

No type of model was notably more successful than the others in raising scores on cognitive conceptual skills. No model type had an overall average positive effect on measures in this domain, which included reading comprehension and problem solving. One model that did have considerable impact on cognitive conceptual skills was the Direct Instruction model. Not one 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 skills. Models that focused on cognitive-conceptual skills were incapable of influencing standardized measures of those skills after four years of instruction.

Models that emphasized basic skills produced better results on tests of self-concept than did other models. On the average, children in models the evaluators classified in the Basic Skills category, performed better on affective measures than did children in Cognitive-Conceptual or Affective models. All models in the Basic Skills category had positive average model effects. The only other model to demonstrate a positive average effect was the University Florida's Parent Education model. In every case, the models that focused on affective development had negative average effects on 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 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 an outcome of skill

Abt evaluators concluded that a Basic Skills model would be preferable if an educator was concerned with teaching skills such as spelling, math computation, language, and word knowledge. ...The Direct Instruction model had an unequivocally higher average effect on scores in the basic skills domain than did any other model.

acquisition. In other words, rather than considering self-concept a necessary prerequisite for learning, they contend that instruction resulting in academic success leads to improved self-concept. The data uphold this view.

With the exception of the Florida Parent Education model, all Cognitive-Conceptual and Affective-Cognitive models had more negative than positive outcomes on measures in the basic skills domain.

It would be a mistake, however, to 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 Basic Skills models, Direct Instruction and Behavior Analysis. In other words, describing the result as a "Basic Skills" effect does not identify the specific instructional variables that lead to significantly better performance in both outcome areas. The fact remains however, that no model classified as "Affective" had a positive average effect on affective measures.

The average effects for nine individual models are represented in Figure 1. The centerline of the figure indicates no difference between students in a Follow Through model and comparison students. Notice that the Direct instruction model is the *only* model to show sizable positive effects on all measures. The majority of models show considerable negative effects (performance below the level of the comparison group) on all measures. These findings clearly show the Direct instruction model to be superior on these measures compared with traditional programs and with other Follow Through models.

The evaluation was not only costly, but controversial. At least three other major reanalyses of the data were independently conducted. None of these analyses show significant disagreement with respect to achievement data. Results of the national evaluation and all subsequent analyses converge on the finding that the highest achievement scores were attained by students in the Direct Instruction model. 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 a clear answer to the question.

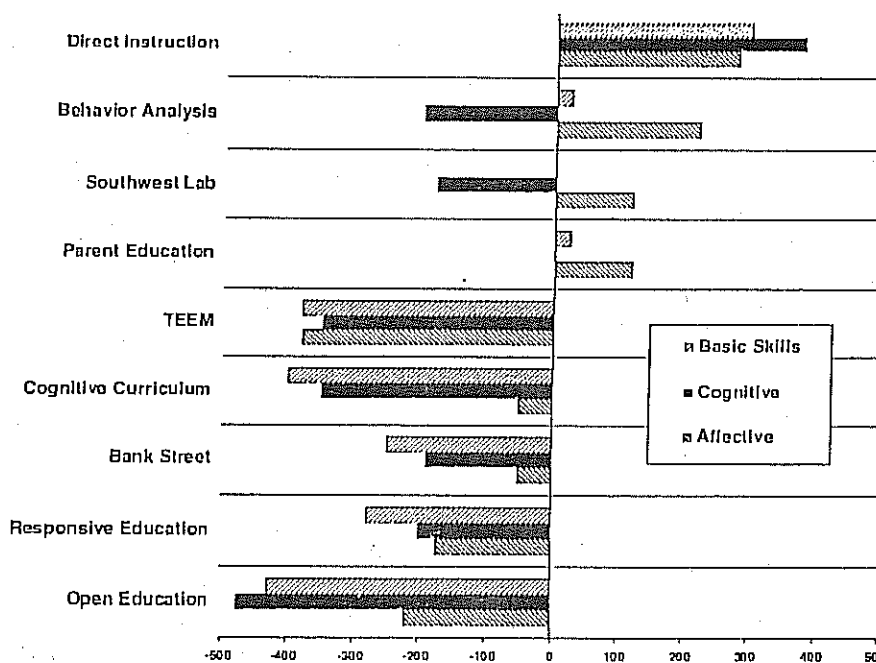


Figure 1

This figure shows the average effects of nine Follow Through models on measures of basic skills (word knowledge, spelling, language, and math computation), cognitive-conceptual skills (reading comprehension, math concepts, and math problem solving), and self-concept. This figure is adapted from Engelmann, S., & Carnine, D. (1982), *Theory of instruction: Principles and application*. New York: Irvington Press.

Not one 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 skills. One model that did have considerable impact on cognitive conceptual skills was the Direct Instruction model.

Dissemination

The purpose of the Follow Through planned variation experiment was to identify effective educational methods. However, there is little utility in identifying effective methods if they are not then made accessible to school districts. The Joint Dissemination Review Panel and the National Diffusion Network were created to validate and disseminate effective educational programs. In 1977, Follow Through sponsors submitted programs to the JDRP. "Effectiveness" was, however, broadly interpreted. For example, according to the JDRP, the positive impact of a program need not be directly related to academic achievement. In addition, a program could be judged effective if it had a positive impact on individuals other than students. As a result, programs that had failed to improve academic achievement in Follow Through were rated as "exemplary and effective." And, once a program was validated, it was packaged and disseminated to schools through the National Diffusion Network.

The JDRP's validation practices did not go unchallenged. According to former Commissioner of

The Direct instruction model is the only model to show sizable positive effects on all measures. The majority of models show considerable negative effects (performance below the level of the comparison group) on all measures. These findings clearly show the Direct instruction model to be superior on these measures compared with traditional programs and with other Follow Through models.

Education, Ernest Boyer, "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, 1984, p. 87). However, commissioner Boyer's concerns could not prevent the widespread dissemination of ineffective instructional approaches. The JDRP apparently felt that to be "fair" it had to represent the multiplicity of methods in education. Not only did this practice make it virtually impossible for school districts to distinguish between effective and ineffective programs, it defeated the very purpose for which the JDRP and NDN were established.

The evaluation was not only costly, but controversial. At least three other major reanalyses of the data were independently conducted. None of these analyses show significant disagreement with respect to achievement data.

Funding Decisions

The effect of the Follow Through evaluation may also be measured by the extent to which the findings influenced decisions about funding. While all Follow Through models received budget cuts over the years, the disbursement of available funds was not based on effectiveness, but on a non-competitive continuation basis. 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 that had not been validated!

Not surprisingly, funding ineffective programs at a higher level did not make them effective. Not one additional program was validated during the following year. Yet the same funding policy continued to be implemented, favoring ineffective programs. It is clear that increased financial support by itself does not lead to increased performance by students. How children are taught is critically important.

The results of the Follow Through evaluation also failed to influence decisions about allocation of federal research funds. Planned variation makes it possible to identify the best performing programs and then subject them to further analyses. Instead the Office of Education and National Institute of

Education agreed to spend 12 million dollars to develop and study new Follow Through approaches with the primary concern being "whether or not an approach can be put in place and maintained, *not with the effectiveness of the approach in improving student outcomes*" Proper and St. Pierre, 1980, p. 8) [emphasis added]. According to Stallings (1975), the Direct Instruction model was not only most effective, it and the Behavior Analysis models were the most easily implemented. If information about implementation was needed, these two models provided a good starting point. The plan that was

According to former Commissioner of Education, Ernest Boyer, "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..."

pursued shows total neglect of the findings of the Follow Through evaluation and astonishing disregard for the academic achievement of students.

Perhaps even more disturbing is the fact that twenty years after the publication of the Follow Through evaluation, there is little evidence that the results have altered educational practices in American classrooms. The majority of schools today use methods that are not unlike the Follow Through models that were least effective (and in some cases were most detrimental). Barriers at all levels of the educational system preclude widespread adoption of the model that was most effective.

The Educational Establishment

The history of Follow Through and its effects constitute a case study of how the educational establishment functions. As in other bureaucracies, it is composed of parochial vested interests that work to either maintain the status quo or to advance a self-serving agenda. As a result, the largest educational experiment in history (costing almost one billion taxpayer dollars) has been effectively prevented from having the impact on daily classroom practices that its results clearly warranted. Let's look at some factors that operate at each level of the educational establishment to influence decisions about teaching methods and materials.

Policymakers. Follow Through demonstrated that public policy is based on public support, not on empirical evidence. Thus, the position that officials adopt with respect to teaching methods is most likely to be congruent with the position of the majority. Because the Direct Instruction model represents a minority view in education, it was not surprising that policymakers failed to take a strong position in support of the Follow Through results.

Although some policymakers may have some formal training in areas of education, they typically rely on input from education professionals when developing and supporting programs. The influence of stakeholders in traditional educational practices can be seen throughout the history of Project Follow Through. Planning committees, advisory boards, and task forces were composed of representatives of universities and research centers. These professionals usually represent educational philosophies that the Follow Through results suggest do not, and cannot, lead 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.

Clearly some professionals with a self-interest have the power to influence educational policy in a direction that will not necessarily lead to improved education. In fact, some social policy analysts assert that in situations where administrators are strongly convinced of the effectiveness of a program, it is likely that an evaluation will be disregarded. This is tragically illustrated in California where policy makers enamored with Whole Language were seemingly incapable of attending to data showing serious declines in students' reading performance, including a national assessment on which California students placed last. By ignoring outcome data, policy makers continue to make educational decisions that negatively impact children. And the most vulnerable learners are those who are most seriously harmed.

While all Follow Through models received budget cuts over the years, the disbursement of available funds was not based on effectiveness, but on a non-competitive continuation basis. ...Not surprisingly, funding ineffective programs at a higher level did not make them effective.

The majority of schools today use methods that are not unlike the Follow Through models that were least effective (and in some cases were most detrimental).

An additional problem is that policymakers frequently rely on information that others provide them. Thus their decisions are often based on incomplete and inaccurate data that reflect not what research has revealed, but the biases of program administrators, and supporters. An Office of Education document that was read at an appropriations meeting claimed that "when contrasting all Follow Through children with their non-Follow Through comparisons... there emerge large differences in achievement, motivation, and intense effects" (U. S. Congress, 1974, p. 2361), a statement leading senators to believe that the Follow Through program as a whole was successful and should be continued. John Evans, OE's Acting Deputy Commissioner for Planning, Budgeting, and Evaluation, explained to Congress that:

...Follow Through is made up of a different set of alternative ways of approaching alternative education, different models, different programs. And the task and central purpose of that program...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...other models are not successful and not effective and not the kinds of things we would want to carry on or continue to fund or support indefinitely (U. S. Congress, 1974, p. 2360).

This example illustrates how reports and interpretation of results may serve as a source of confusion when decision makers are faced with the task of determining the fate of a program.

It is acknowledged that policy makers are more likely to be influenced by social and political contingencies than by empirical data. However, others

may be expected to pay more heed to the findings of major research programs in their field.

Colleges of Education. Project Follow Through was unique because it examined not only instructional programs, but the educational philosophies from which they were developed. 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 and Carnine, 1981) or what Bijou (1977) 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. Whole Language is an example of instruction derived from this philosophy. It is based on the belief that reading develops naturally given sufficient exposure to a print-rich environment.

The second philosophical position is concerned with principles of learning or "changing behavior" (Becker and Carnine, 1981). From this perspective, teaching involves specifying what is to be taught

The history of Follow Through and its effects constitute a case study of how the educational establishment functions. As in other bureaucracies, it is composed of parochial vested interests that work to either maintain the status quo or to advance a self-serving agenda.

and arranging the environment in such a way that the desired change in behavior results.

Although the data from Follow Through support the latter position, the majority of colleges of education espouse a philosophy of cognitive restructuring. Thus, the data from Follow Through fail to support the philosophy that dominates colleges of education. This obviously made it difficult for educators to accept the Follow Through findings and they responded by discrediting the evaluation as well as by voicing specific objections about the Direct Instruction model or questioning the values of the model. For example, educators are fond of accusing direct teaching approaches of ignoring the "whole child" by emphasizing academic achievement at the expense of affective development. The Follow Through data clearly show that no such trade-off occurs. The Direct Instruction model was

Educators are fond of accusing direct teaching approaches of ignoring the "whole child" by emphasizing academic achievement at the expense of affective development. The Follow Through data clearly show that no such trade-off occurs. The Direct Instruction model was more effective than any other model on measures of self-esteem.

more effective than any other model on measures of self-esteem. A second objection is that this Direct instruction is reductionistic and results in only rote learning of non-essential skills. Yet, the data show that students in the Direct Instruction model demonstrated superior performance on measures of complex cognitive skills. In contrast, not a single model that set out to improve these cognitive skills was able to do so.

Although effective methods may be rejected simply because of their philosophical underpinnings, it is possible that they are rejected for more practical reasons as well. If teachers are to become competent in the use of effective teaching methods, teacher

Student failure is more likely to be attributed to deficits within the child or to external factors such as the child's home life, than to ineffective instruction.

training programs must be restructured and those who are responsible for teacher training must themselves become proficient in those methods. Effective restructuring will require changes not only in what is taught, but in how it is taught as well. The training paradigm underlying most teacher training programs has little to recommend it, with students spending the majority of their time listening to lectures about theory and method. Sponsors of Follow Through models found that lectures about teaching had little impact upon actual teaching practices. Training was most successful when it included modeling of the desired behaviors, opportunities for teachers to practice, and feedback about their performance (Bushell, 1978). This has important implications not only for preservice training of teachers,

but for how schools provide inservice training.

Teachers. Probably the biggest obstacle is the fact that the instructional methods a teacher uses are most likely to be those taught during his or her own training. Although it is assumed that teachers have acquired the skills necessary to teach their students, in reality teachers are woefully unprepared. For example, there are currently thousands of teachers in classrooms who do not know how to teach beginning reading, because the professors who "taught" them adhered to a philosophy of "natural growth." One teacher confided to me, "I do not know how to teach reading to someone who doesn't already know how to read"! If our teachers do not, by their own admission, know how to teach, how will our children learn?

Teachers may not seek out empirically validated methods, such as Direct Instruction, because they fail to recognize that their current methods are not effective. Student failure is more likely to be attributed to deficits within the child or to external factors such as the child's home life, than to ineffective instruction. Furthermore, many teachers are not even aware that methods exist that would enable them to be more effective. In many instances, the only information teachers have about Direct Instruction is misinformation. And, even if teacher did know there was a better way to teach, how would they acquire the necessary skills? Surely not by returning to the schools where they received their initial teacher training.

Teachers who are motivated to look for and use effective methods, often run into opposition. For example, if Direct Instruction materials have not been approved for purchase by curriculum committees, teachers will, in effect, be unable to purchase those materials. Even if appropriate materials can be obtained, teachers may be forbidden to use them. In addition, districts often refuse to provide funds for teachers to attend Direct Instruction conferences and training sessions, preferring to send them to receive information about the most current fads.

School Districts. The fact that effective teaching methods are available does not mean that they will be adopted. According to Alan Cohen (personal communication, 1992), "We know how to teach kids,

Few educational programs are based on state-of-the-art programming principles. Worse yet, materials are not field tested to ensure their effectiveness with children.

what we don't know is how to get the public schools to do it!" Because there are no incentives for adopting effective methods or penalties for adopting ineffective ones, the choice of instructional programs will be made based on other factors. One factor that determines whether a particular method will be adopted is how greatly it differs from existing practices. The best candidates for adoption are those most similar to ongoing practices, because they are least disruptive. Stallings and Kaskowitz (1974) described the behavior of teachers in Direct Instruction Follow Through classrooms as "quite atypical of generally practiced classroom behavior" (p. 220). This decreases the probability of adoption because it requires so much change.

Parents and others have been led to accept that the failure of a great many students to learn is due to deficits in the children. The general public has no way of knowing that children's achievements are largely a function of how they are taught.

Financial incentives may also influence adoption decisions. While funding may provide the inducement to adopt an innovation, monitoring is needed to ensure its continued implementation. One way that Follow Through differed from other federally funded programs was that in exchange for funding, particular instructional practices were specified and monitored. This system of supervision resulted in a higher degree of fidelity of implementation of the model than might otherwise be expected. However, schools are generally not organized to provide the level of supervision that Follow Through model sponsors found necessary to ensure fidelity of implementation.

Publishers. Much, perhaps most, of what a teacher does is determined by the materials he or she uses. Yet, those who develop instructional materials typically do not have the skills required to develop effective materials. Few educational programs are based on state-of-the-art programming principles. Worse yet, materials are not field tested to ensure their effectiveness with children. The publishing industry does not initiate the development of instructional materials, but instead reacts to the demands of the educational marketplace. California provides a good illustration of this dependent relationship. In California the state adopts an instructional framework. Criteria for instructional materi-

als are then derived from the framework. Publishers are provided these criteria and busily get to work developing instructional materials that conform to them. They submit their materials during the textbook adoption process and panels evaluate the extent to which the materials correspond to the specified criteria. Noticeably absent from these criteria is any mention of measured effectiveness. Given this process, a program could meet every single criterion and be recommended for adoption, and not be effective in teaching a single child! But, field tests are expensive, and the prevailing contingencies provide absolutely no incentive for publishers to conduct them in order to provide learner verification data because such data are not considered in textbook selection and adoption. (See "Why I sued California, Engelmann, *ADI News*, Winter, 1991).

The Public. Although the public is not typically considered part of the educational establishment, it can be included in this discussion because it supports education. What the public has supported is a system which has continued to neglect effective methods of instruction. Of course, the public's support has been innocent because it is generally unaware of instructional options and their differential effectiveness. Parents and others have been led to accept that the failure of a great many students to learn is due to deficits in the children. The general public has no way of knowing that children's achievements are largely a function of how they are taught. However, this may be changing.

Our educational problems will not be solved until it is recognized that how well students learn is directly related to how well they are taught.

Toward the Future

The American public's dissatisfaction with public education is becoming increasingly clear. The failures of public education have been well publicized. Endless studies and reports call attention to important factors such as improving curricula, increasing teacher salaries, expanding the length of the school day and/or year, and a variety of other changes. Although some of these changes may be necessary, they will not be sufficient to produce the substantial academic improvement that is possible. The critical factor that has been historically ignored is instructional method. Our educational problems will not be solved until it is recognized that how well students learn is directly related to how well they are taught.

Is there any evidence that research is beginning to influence educational policy and practice? Recent events in California may point to progress in that direction. The Report of the California Reading Task Force (1995) stresses effective teaching and recommends that every school and district implement a "reading program that is research based" (p. 3). In February of this year, Assembly Bill 3075 (1996) was introduced in the California State legislature. This bill would amend the minimum requirements for a teaching credential to include satisfactory completion of comprehensive reading instruction "that is research-based" and includes "the study of direct, systematic, explicit phonics." In September of 1995, Governor Wilson signed Assembly Bill 170, referred to as the ABC bill, requiring the State Board of Education to "ensure that the basic instructional materials that it adopts for mathematics and reading...are based on the fundamental skills required by these subjects, including, but not limited to systematic, explicit phonics, spelling, and basic computational skills." It is possible that these documents offer the promise of hope for the future. I will close with the words of Leonardo da Vinci: "Tell me if anything ever was done."

References

- Becker, W. C. & Carnine, D. (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: Erlbaum.
- Bijou, S. W. (1977). Practical implications of an interactional model of child development. *Exceptional Children*, 14, 6-14.
- 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.
- Carnine, D. W. (1983). Government discrimination against effective educational practices. *Proceedings of the Subcommittee on Human Resources Hearing on Follow Through Amendments of 198*, 99-103. Wash. D. C.: U. S. Government Printing Office.
- Carnine, D. W. (1984). The federal commitment to excellence: Do as I say, not as I do. *Educational Leadership*, 4, 87-88.
- Engelmann, S. (1991). Why I sued California. *ADI News*, 10, 4-8.
- Engelmann, S. & Carnine, D. (1982). *Theory of instruction: Principles and applications*. New York: Irvington Publishers.
- Every child a reader: The report of the California reading task force*. (1995) Sacramento: California Department of Education.
- Proper, E. C., & St. Pierre, R. G. (1980). *A search for potential new Follow Through approaches: Executive summary*. Cambridge, Mass.: Abt Associates. (ERIC Document Reproduction Services No. ED 187 809)
- Stallings, J. A., & Kaskowitz, D. H. (1974). *Follow Through classroom observation evaluation (1972-1973)*. Menlo Park, CA: Stanford Research Institute.
- 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, Mass.: Abt Associates.
- U. S. Congress, Senate. (1974, May 29). Hearings before a subcommittee of the Committee on Appropriations.

Our Failure To Follow Through

Billy Tashman

Reprinted from New York Newsday, November 15, 1994, with permission

Project Follow Through, America's longest, costliest and perhaps, most significant study of public school teaching methods quietly concluded this year. The good news is that after 26 years, nearly a billion dollars, and mountains of data, we now know which are the most effective instructional tools. The bad news is that the education world couldn't care less.

Started in 1968, Follow Through was intended to help kids, from kindergarten through the third grade, continue the progress they had made in Head Start. But the Feds also wanted the find out which instructional methods delivered the most bang for the bucks. So they funded 22 vastly different educational programs in 51 school districts with a disproportionate number of poor children. Standardized test results were collected from almost 10,000 Fol-

The good news is that after 26 years, nearly a billion dollars, and mountains of data, we now know which are the most effective instructional tools. The bad news is that the education world couldn't care less.

low Through children, as well as from kids not in the Follow Through program.

Abt Associates in Cambridge, Mass., analyzed the numbers, then issued the verdict. When it came to academic performance, children who participated in the Direct Instruction method blew their peers out of the classroom. More important, later evaluations of 1,000 Direct Instruction graduates showed that they were still ahead of their cohorts in their senior year of high school.

If something works this well, why aren't public schools using it? One reason is that Direct Instruction, at first glance, looks dated. Indeed, teachers who treat their jobs as a cross between stand-up comedy and the Superbowl halftime show might, after peeking into a Direct Instruction classroom, disappear faster than a spare textbook at the Board of Ed.

To make matters worse, these methods owe a lot to the late B. F. Skinner, the Harvard behaviorist some recklessly called a fascist. That's unfortunate and unfair, because Skinner demanded a scientific

approach to classroom instruction, which is lacking from almost every hot reform idea *du jour*.

Direct Instruction stresses basic skills, breaking them down into mini-components. Children learn to read, for example, by learning the sounds of the letters before the letter names. They master each skill before moving onto the next one. Teachers track each student's progress on daily charts. They also track behavior, encouraging good conduct with praise, while ignoring bad behavior for the most part. In short, if you can't measure it, you probably shouldn't teach it. This kind of micro-management is almost unheard of in most classrooms.

But Direct Instruction's most controversial feature is a script from which teachers conduct lessons. Picture this: A first-grade teacher, reading from her script, makes the "m" sound. The pupils respond in unison. After a word of praise, the teacher, prompted by her script, tells them to repeat the sound.

This may sound a bit like a "Road to Wellville" approach to education, but Direct Instruction has had stunning success at scores of schools. One of the original sites in the early '70s was P.S. 77 in the South Bronx. After five years, DI "significantly raised the reading, writing and arithmetic performance and scores of the participating children," said one report. Federal budget cuts eventually gutted the program but, interestingly, P.S. 77 old-timers still cling lovingly to the teaching methods.

It may come as a shock to the layperson, but school policymakers haven't adopted Direct Instruction because they have an aversion to scientific research. Educators throw their weight behind the latest fad, then refuse to abandon it when it doesn't work. In fact, the federal oversight panel for Follow Through cut the Direct Instruction program even as it continued other models that were spectacular flops. Eschewing basic skills, the failed programs tried to teach kids how to learn on their own, or tried to raise students' self-esteem (both categories, by the way, in which Direct Instruction students excelled). In these failed programs, students had even lower reading and math scores than the control groups that had no Follow Through program. Yet these failed programs have spread through America like fire through dry corn.

Follow Through demonstrated that scientific research and the classroom are still strangers to one another. Until they join forces, American schoolchildren will continue to receive a second-class education.

Videotapes on the Direct Instruction Model

Keynotes from the 1995 Conference—2 hours. Titles and speakers include: Anita Archer, Professor Emeritus, San Diego State University, speaking on "The Time Is Now" (An overview of key features of DI); Rob Horner, Professor, University of Oregon, speaking on "Effective Instruction for All Learners;" Zig Engelmann, Professor, University of Oregon, speaking on "Truth or Consequences."

Price: \$25.00

Keynote Presentations from the 1994 20th Anniversary Conference—2 hours. Titles and speakers include: Jean Osborn, Associate Director for the Center for the Study of Reading, University of Illinois, speaking on "Direct Instruction: Past, Present & Future;" Sara Tarver, professor, University of Wisconsin-Madison, speaking on "I have a Dream That Someday We Will Teach All Children;" Zig Engelmann, Professor, University of Oregon, speaking on "So Who Needs Standards?"

Price: \$25.00

An Evening of Tribute to Siegfried Engelmann—2.5 hours. On July 26, 1995, 400 of Zig Engelmann's friends, admirers, colleagues, and protégés assembled to pay tribute to the "Father of Direct Instruction." The Tribute tape features Carl Bereiter, Wes Becker, Barbara Bateman, Cookie Bruner, Doug Carnine, and Jean Osborn—the pioneers of Direct Instruction—and many other program authors, paying tribute to Zig.

Price: \$25.00

Challenge of the 90's: Higher-Order thinking—45 minutes, 1990. Overview and rationale for Direct Instruction strategies. Includes home-video footage and Follow Through.

Price: \$10.00 (includes copying costs only).

Follow Through: A Bridge to the Future—22 minutes, video, 1992. Direct Instruction Dissemination Center, Wesley Elementary School in Houston, Texas, demonstrates approach. Principal, Thaddeus Lott, and teachers are interviewed and classroom footage is shown. Created by Houston Independent School District in collaborative partnership with Project Follow Through.

Price: \$10.00 (includes copying costs only).

Where It All Started—45 minutes. Zig teaching kindergarten children for the Engelmann-Bereiter pre-school in the 60's. These minority children demonstrate mathematical understanding far beyond normal developmental expectations. This acceleration came through expert teaching from the man who is now regarded as the "Father of Direct Instruction," Zig Engelmann.

Price: \$10.00 (includes copying costs only).

Direct Instruction—black and white, 1 hour, 1978. Overview and rationale for Direct Instruction compiled by Haddox for University of Oregon College of Education from footage of Project Follow Through and Eugene Classrooms.

Price: \$10.00 (includes copying costs only).

Corrective Reading: Decoding B1, B2, C—4 hours, 38 minutes + practice time. Pilot video training tape that includes an overview of the Corrective Series, placement procedures, training and practice on each part of a decoding lesson, information on classroom management / reinforcement and demonstrations of lessons (off-camera responses).

Price: \$25.00 per tape (includes copying costs only).

Order from ADI
VISA of Mastercard accepted
Call 1-800-995-2464
ADI, PO Box 10252, Eugene, OR 97440

READING MASTERY / FAST CYCLE TRAINING VIDEO SERIES

Reading Mastery I – 10 videotapes
Reading Mastery II – 5 videotapes
NEW! Reading Mastery III videotape

The first five tapes in both the Level I and Level II series present intensive pre-service training on basic Direct Instruction teaching techniques and classroom management strategies used in READING MASTERY and the equivalent lessons in FAST CYCLE. **Rationale** is explained. **Critical techniques** are presented and **demonstrated**. Participants are led through and then **practice** techniques. Classroom teaching demonstrations with students are shown. The remaining tapes in each series are designed to be used throughout the school year. The tapes are divided into segments which present critical teaching techniques for a set of upcoming lessons. These tapes can serve as a critical part of in-service training.

Level III is presented on one videotape with the same high-quality training features as described above for the first two levels. A training manual is included with each video training package.

Name: _____

Institution: _____

Mailing Address: _____

Telephone: _____ Fax: _____

Purchase Price

_____ set RM/ FC training tapes 1-10 with guide @ \$229.00

_____ set RMII/ FC training tapes 1-5 with guide @ \$129.00

_____ combined package—RM I and II tapes @ \$299.00

_____ RMIII training tape and guide @ \$30.00

Shipping and Handling

_____ shipping costs—\$20/one set, \$15/each additional set

Total \$ _____

Make check or purchase order payable to D.I.T.V.

D.I. Training Videos

PO Box 10459

Eugene, OR 97440

(541) 485-1163

FAX (541) 683-7543

ADI MATERIALS PRICE LIST

Theory of Instruction (1991)

by Siegfried Engelmann & Douglas Carnine

Membership Price: \$32.00

List Price: \$40.00

The Surefire Way to Better Spelling (1993)

by Robert C. Dixon

Membership Price: \$8.75

List Price: \$12.00

Teach Your Child to Read in 100 Easy Lessons (1983)

by Siegfried Engelmann, Phyllis Haddox, & Elaine Bruner

Membership Price: \$14.95

List Price: \$17.95

Teacher Monitoring Program (1992)

by Colin Bird, Elizabeth Fitzgerald, & Margaret Fitzgerald

Membership Price: \$15.00

List Price: \$15.00

Structuring Classrooms for Academic Success (1983)

by Stan Paine, J. Radicchi, L. Rosellini, L. Deutchman, & C. Darch

Membership Price: \$11.00

List Price: \$14.00

War Against the Schools' Academic Child Abuse (1992)

by Siegfried Engelmann

Membership Price: \$14.95

List Price: \$17.95

Turning Our School Around: Seven Commonsense

Steps to School Improvement (1995)

by Phyllis Anderson Wilken

Membership Price: \$15.95

List Price: \$19.95

Postage & Handling:

If your order is:

P & H is:

\$0.00 to \$20.99

\$4.00

\$21.00 to \$40.99

\$5.50

\$41.00 to \$60.99

\$7.00

\$61.00 to \$80.99

\$8.50

\$81.00 or more

\$10.00

Subtotal

P & H

Outside the continental U.S., add \$3 more

ADI Membership Dues

Total

(U.S. Funds)

Check enclosed. ☐ (Make checks payable to Association for Direct Instruction.)

Bill my credit card. ☐ Circle one: VISA Mastercard

Number:

Expiration date:

Signature:

Name (please print):

Address:

You may also phone in your order with VISA or Mastercard.
1-800-995-2464

Send to:
ADI
PO Box 10252
Eugene, OR
97440

Recommended Resources

School's Out: The Catastrophe in Public Education and What We Can Do About It (1993) by Andrew Nikiforuk.

ISBN: 0-921912-48-X

Price: \$19.95 from Macfarlane Walter & Ross
37A Hazelton Avenue
Toronto, CA M5R 2E3

Ask for it at your local bookstore.

Beginning to Read: Thinking and Learning About Print (1990) by Marilyn Jager Adams (A summary by the Center on Reading).

Price: \$7.50

Mail orders to: Center for the Study of Reading
University of Illinois
51 Gerty Cr.
Champaign, IL 61820

Direct Instruction Reading (Revised, 1990)

by Douglas Carnine, Jerry Silbert, & Ed Kameenui.

Price: \$40.00

Order from: MacMillan Publishing
1-800-257-5755
ISBN: 0-675-21014-3

Antisocial Behavior in Schools: Strategies and Best Practices (1995) by Hill Walker, Geoff Colvin, & Elizabeth Ramsey.

Price: \$28.70

Order from: Brooks/Cole Publishing Co.
1-408-373-0728 (ext 137)
Fax: 1-408-375-6414
Email:

adrienne_carter@brookscole.com

(Complimentary copies sent for review for college course. Send request on letterhead.)

Failing Grades (Video) and Annotated Bibliography (1993) featuring Joe Freedman, M.D. & Mark Holmes, Ph.D.

Price: \$17.95

Order from: Society for Advancing Research
c/o VICOM Limited
11603-165 Street
Edmonton, Alberta
CANADA T5M 3Z1

If Learning Is So Natural, Why Am I Going To School? (1994) by Andrew Nikiforuk.

Price: \$16.99 from Penguin

ISBN: 0-14-02.4264-3

Ask for it at your local bookstore.

Becoming a Nation of Readers (1985)
The Report of the Commission on Reading.

Price: \$7.50

Mail orders to: Center for the Study of Reading
University of Illinois
51 Gerty Cr.
Champaign, IL 61820

Direct Instruction Mathematics (Revised, 1990) by Jerry Silbert, Douglas Carnine, & Marcy Stein.

Price: \$40.00

Order from: MacMillan Publishing
1-800-257-5755
ISBN: 0-675-21208-1

Interventions for Achievement and Behavior Problems (1991) by 74 contributors, edited by Gary Stoner, Mark Shinn, & Hill Walker.

Price: \$52.00

Order from:
National Association of School Psychologists
8455 Colesville Road, Suite 1000
Silver Spring, MD
ISBN: 0-932955-15-0

Higher Order Thinking: Designing Curriculum for Mainstreamed Students (1992) edited by Douglas Carnine and Edward J. Kameenui.

Price: \$24.00 (prepaid orders postage-free)

Order #5199 from: PRO-ED
8700 Shoal Creek Boulevard
Austin, TX 78758-9965
FAX: 512-451-8542
ISBN 0-89079-557-6

Effective School Practices
ADI
P.O. Box 10252
Eugene, OR 97440

Non-Profit
Organization
U.S. Postage
PAID
Permit No. 122
Eugene, OR

Dear friend:

This is a complimentary copy of our quarterly publication *EFFECTIVE SCHOOL PRACTICES*. I hope you find the contents helpful and informative. If you have not already joined the *Association for Direct Instruction*, consider joining today.

The mission of the *Association for Direct Instruction* is to improve education. Currently, a major obstacle to improved education is old fads that are recycled as "new reforms," in spite of their failure in the past. Future ADI publications will define instruction that represents the state-of-the-art in research-based practices—instruction that emphasizes thoughtful balance rather than faddish extremes.

Join us! Help counteract faddism in education.
Our membership fee is only \$20.00.

With your membership you will receive a subscription to *EFFECTIVE SCHOOL PRACTICES*, and discounts on publications and on staff development workshops. We are also planning a series of monographs on specialized topics. ADI members will receive free copies of these special ADI monographs at the time of printing.

Bonnie Grossen,
Editor, *Effective School Practices*