

Follow Through Evaluation

DI, UNDISPUTED WINNER

The formal evaluation of Follow Through sponsors and Follow Through's overall performance came out in April of 1977. At that time scientists had recently discovered the cause of Legionnaire's disease; Jimmy Carter had become President and had pardoned Vietnam War draft evaders. The worst air collision to that time had occurred when two planes collided over the Canary Islands killing 583 people.

The Follow Through evaluation did not make headlines because it had not been officially announced or interpreted by the Office of Education. The design of the evaluation was careful. To assure that the analysis was experimentally pristine, two organizations dealt with data. Stanford Research Institute collected the data; Abt Associates, on the other side of the continent, analyzed it. The evaluation cost \$30 million.

At the time of the evaluation, Bob Egbert was no longer the national director of Follow Through. He had been succeeded by Rosemary Wilson, who shared neither his vision nor his conception that if Follow Through permitted sponsors to implement in a conducive context, the evaluation would clearly identify winners and losers.

The analysis centered around the third-grade performance of the later cohorts that went through the various Follow Through programs. 40,000 third-graders were tested. Part of the evaluation did not involve all of our sites, only most of the sites that started in kindergarten. Grand Rapids was one of the sites included in this part of the evaluation, although we had not worked with the site for three years and had received no money from

national Follow Through for sponsoring the site. Yet, the analysis treated Grand Rapids as if it were one of our sites. Even so, we were confident that the data would show that we had won the horse race.

This confidence was in defiance of the educational community's consensus that there would be no clear winners or losers. Although sponsors were not permitted to publish data that could be construed as comparing performance of different models, National Follow Through analyzed the data and reported comparative data to sponsors as early as 1973, when a conference in Brookings, Oregon presented results from 1971. The book *Planned Variation in Education—Should We Give Up or Try Harder?* drew the conclusion, "It already seems highly doubtful, however, that the results will provide clear-cut indications that one model is best."

This conclusion was based on the 1970–71 cohort, but there was data on the 1971–72 cohort, which generated a far different picture. Kansas and our model were far ahead of the others. Also, the director of Follow Through research, Gary McDaniels, wrote, "Several sponsors looked very strong after the first year, while others did not. The strongest were those that emphasized short-term achievement effects [in other words, Kansas and us]."

The first published reports on the Follow Through performance were based on an analysis of sponsors conducted by Stallings and Kaskowitz, which appeared in *Behavior Today* in 1975. After making extensive observations of the various sponsors' classrooms, Stallings concluded that there were different winners that corresponded to different program emphases. According to Stallings's calculations, those approaches that focused on reading and spent more time on reading had better reading performance (DI and Behavior Analysis). The main problem with this conclusion was that we did not spend more time teaching reading than most of the models. In fact, we probably spent less than half the time provided by the Bank Street model and several others.

Stallings also concluded that different programs were creating children who were different in problem solving, responsibility, question asking, and cooperation. She concluded that children in High Scope and Open Education were high in these traits. She wrote: "Cooperation was marked in classrooms where a wide variety of activities occurred throughout the day and where children would explore and choose their groups." The problem was that she defined cooperation in terms of the activities. If children spent more time in activities that apparently involved cooperation, Stallings concluded that they were "more cooperative."

She also defined responsibility in terms of activities—a child or group engaged in any task without an adult. The definition has nothing to do with the amount of responsibility children learn. (Of course she assumed that if children spend time unsupervised, they must be learning more about

responsibility.) If institutionalized children had been included in the evaluation, they probably would have had “responsibility” scores even higher than Open Education or High Scope because they often have no supervision.

During the period before the Abt Report came out, I was not concerned with what the analysts said about the data. I didn’t have either time or interest to debate whether the fat lady was singing yet.

The fourth volume of the Abt Report presented data on Follow Through sponsors. It came out in 1977 and left no doubt about whether the fat lady had sung. The volume provided arias involving winners and losers, based on performance data. The report confirmed what we knew all along. No other model was close to ours in sophistication.

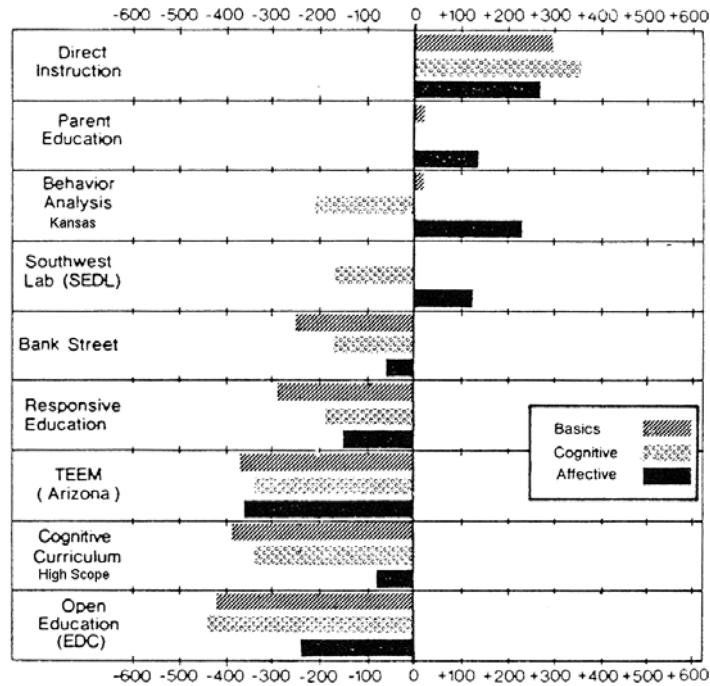
The achievement-test data and that of other tests were analyzed two ways, “adjusted” and “unadjusted.” The adjusted data were expressed as positive or negative outcomes. If a particular site had a score that was a standard number of points higher than the other sites, the site received a plus (+). If the site had a score that was a standard number of points lower than the other sites, the site received a minus (–). If the site was somewhere between a + and –, the difference was considered educationally insignificant.

For analyzing performance of sponsors, Abt threw out performance comparisons if the Follow Through site and the comparison groups differed by more than 50 percent on their entry scores. Several comparisons involving East St. Louis were thrown out, which was unfortunate because East St. Louis children were initially more than 50 percent lower than the children in the comparison groups but still outperformed them by enough to earn a +.

There were other ways the analysis was bent to be unkind to DI, including the way some of the data were “interpreted.” Even so, the numbers didn’t lie.

The evaluation had three categories: basic skills, cognitive (higher-order thinking) skills, and affective responses. The graph on page 226 shows the outcomes for the nine major sponsors.

**Number of Significant Outcomes for Basic Skills (B),
Cognitive Skills (C), and Affective Measures (A)**



Index of Significant Outcomes

The basic skills consisted of those things that could be taught by rote—spelling, word identification, math facts and computation, punctuation, capitalization, and word usage. DI was first of all sponsors in basic skills. Our average score was +297 (which means that we had a considerably larger number of significantly positive outcomes than the Title I comparison students). Only two other sponsors had a positive average. The remaining models scored deep in the negative numbers, which means they were soundly outperformed by children of the same demographic strata who did not go through Follow Through. The sites that Stallings glorified did poorly in this category. High Scope was -389 and Open Education was -433 (far fewer significantly negative outcomes than the Title I comparisons recorded). According to the Stallings predictions, this outcome might be expected for basic skills.

DI was not expected to outperform the other models on “cognitive” skills, which require higher-order thinking, or on measures of “responsibility.” Cognitive skills were assumed to be those that could not be presented as rote, but required some form of process or “scaffolding” of one skill on another to draw a conclusion or figure out the answer. In reading, children were tested on main ideas, word meaning based on context, and inferences. Math problem-solving and math concepts evaluated children’s higher-order skills in math.

Not only was the DI model number one on these cognitive skills; it was the only model that had positive scores for all three higher-order categories: reading, math concepts, and math problem-solving. DI had a higher average score on the cognitive skills (+354) than it did for the basic skills (+297). No other model had an average score in the positive numbers for cognitive skills. Cognitive Curriculum (High Scope) and Open Education performed in the negative numbers, at -333 and -450.

On the affective measures, which included a battery of tests that evaluated children’s sense of responsibility and self-esteem, our model was first, followed by Kansas. The models that stressed affective development performed even below the Title I average.

One of the affective tests described positive achievement experiences and negative experiences. DI children saw themselves as being more responsible for outcomes than children in any other model. On the test that assessed children’s feelings about how they think other people view them and how they feel about school, DI children had the highest scores.

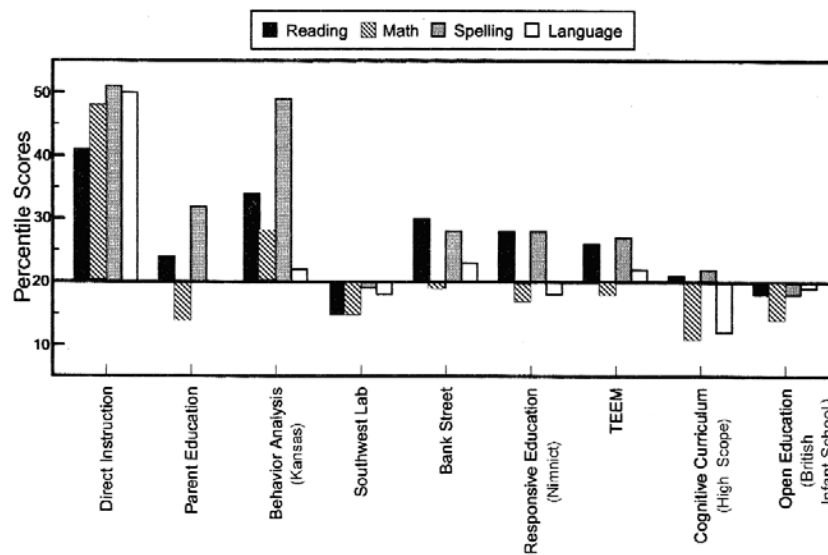
Note that DI was over 250 points above the Title I norm and Open Education was over 200 points below the norm. The Abt Report observed that the high performance of children in our model was unexpected because we did not describe affective outcomes as an objective. The reason was that we assume that children are fundamentally logical. If we do our job of providing them with experiences that show they are smart, they will conclude that they are smart. If they experience success in school that can also be measured in the neighborhood, those experiences serve as fuel for the conclusion that students are competent. At the time of the evaluation, I had heard more than 100 stories of our children helping older siblings learn to read or do homework. The children knew that they could do things the average kid on the street could not do.

The rest of the Abt results were expressed as percentiles. The performance of an average student taking an achievement test is the 50th percentile. The goal for Follow Through had been to achieve the 50th percentile with at-risk children. The average percentile for Title I students was the 20th percentile—less than half that of the average student. The 20th percentile was used as a measure of whether a model produced a posi-

tive or negative effect. The farther above the 20th percentile a model is, the better it performs.

The figure below shows the performance of the nine major sponsors in total reading, total math, spelling, and language.

Percentile Comparisons for the 9 Major Follow Through Sponsors



The horizontal line indicates the 20th percentile. As the figure shows, the competition is closest for reading. All but three of the sponsors scored at or above the 20th percentile. For math, only two models were above the 20th percentile, ours and Kansas (Behavioral Analysis). For language, our model was the only one above the 23rd percentile. High Scope (Cognitive Curriculum) had the lowest scores of all sponsors in math and language. Obviously, this data makes a mockery out of Stallings's notion that an untrained observer could make a few observations of classrooms, classify the activities, and draw any kind of valid conclusion about how successful each program was in inducing cognitive skills.

There's more: Not only were we first in adjusted scores and first in percentile scores for basic skills, cognitive skills, and perceptions children had of themselves, we were first in spelling, first with sites that had a Headstart preschool, first in sites that started in K, and first in sites that started in grade 1. Our third-graders who went through only three years

(grades 1–3) were, on average, over a year ahead of children in other models who went through four years—grades K–3. We were first with Native Americans, first with non-English speakers, first in rural areas, first in urban areas, first with Whites, first with Blacks, first with the lowest disadvantaged children, and first with high performers.

From a historical perspective the performance of our children set important precedents.

1. For the first time in the history of compensatory education, DI showed that long-range, stable, replicable, and highly positive results were possible with at-risk children of different types and in different settings. Previously, “the exemplary program” was a phantom, something that was observed now but not a year from now. Most were a function of fortuitous happenings, a measurement artifact, or a hoax.
2. DI showed that relatively strong performance outcomes are achievable in all subject areas (not just reading) if the program is designed to effectively address the content issues of these areas. Also, this instruction created lively, smart children who had confidence in their abilities.
3. The performance of all the Follow Through children (but particularly DI children) clearly debunked all the myths about DI. DI did not turn children off or turn them into robots. DI children were smart and they knew it.
4. DI outcomes also debunked the myth that different programs are appropriate for children with different learning styles. The DI results were achieved with the same programs for all children, not one approach for higher performers and another for lower performers, or one for non-English speakers and another for English speakers. The programs were designed for any child who had the skills needed to perform at the beginning of a particular program. If the child is able to master the first lesson, she has the skills needed to master the next lesson and all subsequent lessons. The only variable is the rate at which children proceed through the lessons. That rate is based solely on the performance of the children. If it takes more repetition to achieve mastery, we provide more repetition, without prejudice.
5. The enormous discrepancies in performance between our model and all the others implies that we knew something they didn’t know

about instructing the full range of children in the full range of academic skills. We did not buy into the current labels, explanations, or assumptions about learning and performance. The results suggest that our interpretation was right, and that the philosophies of the cognitive and affective models did not translate into effective instruction.

6. The performance of the sponsors clearly debunked the notion that greater funding would produce positive results. All sponsors had the same amount of funding, which was more than a Title I program received. DI performed well in this context; however, the same level of funding did not result in significant improvement for the other models. For all programs there were comprehensive services, which included breakfast, lunch, medical, and dental care, and social services. In this context, the only reasonable cause for the failure of other models was that they used inferior programs and techniques.
7. The Direct Instruction model was the only one that was effective with extremely low performers. We showed that these children could uniformly be taught to read by the end of kindergarten and read pretty well by the end of first grade. Performance of this magnitude and consistency had never been demonstrated in the schools before Follow Through.
8. The relative uniformity of the DI sites implies that DI was better able to make the typical failed teacher successful. Teachers who had not been able to teach children to higher levels of performance were able to do it with our program.
9. Probably most important, the outcome showed that our focus on the moment-to-moment interactions between teachers and children was correct. Most of the other models viewed the problems of instruction in terms of broad interactions between teachers and children, not in terms of specific information delivered in moment-to-moment interactions.

We were not into celebrating, but after work on the day Volume 4 arrived, we had a little party, a couple of beers, and several rounds of congratulations to trainers, managers, and consultants who worked in the trenches to achieve this outcome. Wes reminded us that the game plan was for the winners to be widely disseminated. So we needed to think about tooling up to work with a far greater number of places. We needed to

convert some of our trainers to project managers, recruit some of the super-teachers from our Follow Through sites, and get ready to work with lots of Title I programs. Several of our project managers were skeptical about this degree of acceptance, but the rest of us felt that there would be a payoff for the last nine years of work.

RECONSTRUCTING HISTORY AND LOGIC

With the Abt data published, the moratorium on comparative studies was lifted. We promptly prepared a long article for the *Harvard Educational Review*, “Teaching Reading and Language to the Disadvantaged—What We Have Learned From Research,” which came out in 1977 and gave overviews of our approach and programs, our training, and our results.

We anticipated that the article would stimulate great interest. Instead, there was almost no response—no revelations reported by readers who realized that the practices they espoused had led to unnecessary failure or revelations that DI presented a better way to solve problems that had been haunting school districts since the Coleman Report. There were no frantic phone calls from people wanting to learn more about DI, nor calls from reporters asking about the astonishing results. Instead there was a handful of responses, and most were not positive but raised carping issues about the design of the study or the problems associated with accurately measuring cognitive outcomes. Those who carped had earlier accepted the idea that achievement tests documented the performance problems of at-risk students. Yet, when the same kind of achievement-test data showed that their favored programs produced children who failed as miserably as children summarized by the Coleman Report, they rejected the study.

THE GLASS HOUSE

We later discovered that the effort to trivialize Follow Through data had begun before Abt 4 had been released. The effort was initiated by the Ford Foundation, which had been supporting failed educational programs. 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 a third-party evaluation of Follow Through results. Ernest House was project director. He assembled a panel of professionals with national reputations in their fields—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. This assemblage judged Follow Through data.

The main purpose of the critique was to prevent the Follow Through evaluation results from influencing education policy. The panel's report asserted that it was inappropriate to ask, "Which model works best?" Rather it should consider such other questions as "What makes the models work?" or "How can one make the models work better?"

Glass wrote another report for the National Institute of Education (NIE), which argued that it was not sound policy for NIE to disseminate the results of the FT evaluations, even though the data collection and analysis had cost over 30 million dollars. Here's that part of the 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.

Again, this position is curious for one who apparently believed that data of the same type collected in the FT evaluation earlier documented the problem. Why would the data be adequate to document the problem but not appropriate for documenting outcomes of different approaches that address the problem?

The suggestion that case studies would enable informed choice is not very thoughtful. Qualitative studies work only if they are carefully underpinned with rules about quantities. I'm sure that if the game was for each sponsor to compile descriptions from their high-performing classrooms, DI would have a larger number of success stories than the other sponsors.

Unless there are some number assumptions—like, "How consistently do positive case histories occur?"—The data is useless. Making it even more useless is the depth of description that would be needed to enable an "informed choice." I would guess that each study would require many pages. It would be far more confusing to try to extract information about what works best from these documents than from a few tables that summarize the performance data. In fact, the suggestion for using ethnographic

studies was probably intended to make it impossible for readers to find out what worked best. Who would want to wade through possibly thousands of pages of “case histories” to distill a conclusion about which model was more effective?

Glass even argued that all evaluations involving measurable events and data are invalid. “The deficiencies of quantitative, experimental evaluation approaches are so thorough and irreparable as to disqualify their use.” What is surprising about this statement is how anybody at NIE could have read it and not concluded that the author was loony. The Follow Through experiment was a teaching experiment involving not a few minutes in the lab, but nine years of cohorts in which students passed through four grades in actual classrooms. This study had huge numbers. Also, the evaluation tools that documented performance of students already included ethnographic descriptions of models (prepared by Nero and Associates).

Another argument that Glass presented involved the “audience.”

The audience for Follow Through 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.

There are two major problems with this assertion:

1. The Follow Through study was not designed for teachers but for decision makers—school districts, state and federal departments of education—who serve as gatekeepers for what teachers do. The Coleman Report did not result in individual teachers organizing carpools to schlep children from the inner-city to the suburbs. The Follow Through study was not founded on the assumption that teachers enjoyed some kind of democratic world in which every teacher was able to make independent decisions about what and how to teach. Teachers are not decision makers on policy. Policy makers and district officials are. They would be far better informed by the Follow Through results than by any other single data source because only Follow Through provided extensive comparative data of different approaches.
2. Glass could not have seriously believed that even district-level decision makers would read Abt 4. They wouldn't. Glass appealed to NIE because he was concerned about what NIE would say about Follow Through. The final NIE word would make a lot of news and create great interest. In effect, what NIE would say about the program

would become the truth about it. People, press, and historians would be greatly influenced by NIE's stance.

Also, if NIE followed Glass's recommendations, there would be no challenge to the current order of things in education. The Ford Foundation would save face and wouldn't be labeled as a corporate fool for funding foolish programs for years. People in teacher colleges and district administrators would be able to keep their prejudices about children, learning, and teachers. The publishers of elementary-grade instructional material would be happy because no tidal wave would sink sales of instructional material, and school districts would not have to face uncomfortable issues of overhauling both their belief systems and their machinery. College professors could continue to espouse developmental theories and discovery practices as they decry programs that would "divest teachers of their individuality and creativity."

With a statement that downplayed the Follow Through data, Open Education and High Scope, Bank Street College, and Nimmick's program would not be cast as losers, because ethnographic studies might feature one of their "good" sites. All the primitive but well-greased machinery on all levels from state departments of education to classrooms would remain solidly in place, with no challenge. Of course, somewhere in this political milieu were millions of kids whose lives would be greatly influenced by NIE's decision. But as the anti-number philosophy suggests, who was counting?

In 1978, House and Glass published an article in the *Harvard Educational Review*, "No simple answer: A critique of the Follow-Through evaluation." Unlike Wes's earlier article, this one created quite a stir. A shortened version appeared in *Educational Leadership* in 1978.

Although Gene Glass had been president of the American Educational Research Association, the flaws in the arguments the article presented were so conspicuous that they should have been obvious to the man on the street. The article presented two main arguments to discredit the Office of Education evaluation. The main argument was based on a simple value judgment: sponsors should not be compared. Therefore, the Abt focus on the performance of individual sponsors was inappropriate.

House and Glass contended that the evaluation was actually designed to show how *the aggregate of models* performed, not what individual sponsors achieved. The aggregate failed; therefore, the most definitive statement about Follow Through would simply be Project Follow Through failed. In other words, the average of Follow Through students was no higher than those of comparable Title I students. Therefore, Follow Through failed, which means that every sponsor failed. The question of whether individual sponsors actually failed was not considered relevant because it's bad form to compare sponsors.

What seems most curious about this argument is how House and Glass could conclude that Follow Through as a whole failed. That judgment involves a comparison of programs, Follow Through and Title I. Why is it that sponsors can't be compared but larger programs like Follow Through can?

Even more puzzling was how House and Glass could make the obvious distortion that Follow Through was never intended to evaluate the performance of different sponsors. Certainly, this assertion would be contradicted by Follow Through, unless there was collusion between several agencies, including the Office of Education.

The second House-Glass argument was that no approach was successful with all its sites. House and Glass pointed out that there was variability among each sponsor's sites, some performing well and others poorly. Therefore, no single sponsor should be identified as being "successful."

Again, if House and Glass argued that data could not be used to compare sponsors, by what ground rules were they able to compare sponsors with respect to their variability? Logically, Glass and House would have to throw out this argument or reveal themselves as cherry pickers who used comparative data when they needed it and rejected it when it worked against them.

The variability argument was particularly incredible because it was presented by professionals who were supposed to be experts in experimental designs. The most elementary fact about populations is that every observable population of anything has variation. In fact, populations vary across every measurable feature—the size of the lobes on black oak leaves, the shape of snowflakes, the age of computers. So it would be insane to throw out data simply because it shows that there is variability, particularly in this case because only one DI site varied greatly from the seventeen others and only that site performed poorly. Grand Rapids had third graders performing a year lower than third graders in our other sites. The only possible evidence that House and Glass had about the "failure of DI" or the variability was based on one site that openly rejected the model's provisions and had no contact with the sponsor in more than three years. Furthermore, all the higher-echelon bureaucrats in Follow Through and in NIE knew that we hadn't worked with Grand Rapids for years.

SAD SONG OF THE REAL FAT LADY

The official statement that NIE issued was consistent with the recommendations by Glass and House: Project Follow Through did not significantly improve the performance of disadvantaged students over students in extant Title I programs. There was not a word about winners, losers, or

about the performance of individual sponsors, just a flat statement that Project Follow Through—an aggregate—failed.

Functionally, this decision showed the priorities of the educational system. It was more palatable for educators to accept that their favored approach failed than it was to admit that an approach in disfavor succeeded. The educators' feelings and prejudices were functionally more important to them than evidence that there was a successful method for teaching at-risk children. Stated differently, these people showed that their beliefs were more important than the millions of failed children who could benefit from effective instruction. Make no mistake, they would not have gone through the various machinations they created if they believed their own rhetoric about how important it is to serve at-risk children.

In the end, sites from all the Follow Through models including High Scope and the Open Classroom were "validated." So the status quo was maintained; the models that had horrible results would remain in good standing; all educational myths were perpetuated. If policy makers wanted to believe in instructional models based on student choice, extensive parent involvement, or discovery learning, they wouldn't have to face the pesky problem of how to support this notion with data. Their collective conscience was clear because all these approaches had been "validated." Someone receiving this information would assume that validated means that the validated approach was replicable and sound.

PAPER TRAIL

The master plan for Follow Through and how information about Follow Through would be disseminated to other schools and agencies was complicated.

The switch of emphasis from sponsors to individual school programs had begun as soon as Abt 4 came out in 1977. The Office of Education established the Joint Dissemination and Review Committee. Its purpose was to screen Follow Through schools that applied for funds to disseminate information about the school's successful program. The review committee scrutinized schools that applied through forms, letters, and interviews. Those schools that made it over these hurdles were "validated." Whether they were High Scope schools, Open Education schools, or DI sites. All received the same size validation.

Not all of our sites that applied for funds made the cut. At least four of them received a rating of B and at least two a rating of C, even though they had excellent performance data. It seemed obvious that there was a conscious effort to keep DI from having more representation than some of

the other major models.

Once validated, a school would become a member of the National Diffusion Network (NDN), which consisted of 200 programs, each receiving funds to promote itself to other schools. Of the 200 dissemination schools, only 21 were Follow Through schools, and only 3 were Direct Instruction schools—Flint, Dayton, and East St. Louis. Most of the 200 schools came from a poor list of “effective programs” compiled by the Far West Regional Educational Lab. Very few of these schools actually had data of effectiveness, but neither did at least 14 of the 21 Follow Through sites that were now incorporated into the network.

The National Diffusion Network was possibly well named because it did a good job of diffusing, in the sense of making the effect thin. Instead of being 3 out of the 21 programs to be disseminated, DI was now 3 out of 200, less than 2 percent of the total. And as usual, all “validated sites” had the same status.

Individual schools were eligible to disseminate, but sponsors weren't. The Joint Dissemination and Review Committee ruled that only *schools* could apply for validation, not models.

When we received this news, I thought Wes would go into apoplexy. I had never seen him that angry. I was not a portrait of happiness, but Wes exploded. He quickly recovered and in less than an hour was on the phone, trying to contact senators, representatives, and others who might have some influence.

In October of 1977, Wes and I wrote a letter to Rosemary Wilson protesting Follow Through's position about dissemination. We wrote again in November, after it became apparent that nothing would change. That letter appears in whole, followed by her response. Our letter iterates some of the points I have covered.



UNIVERSITY OF OREGON

College of Education

Department of Special Education

FOLLOW THROUGH PROJECT—"So They Shall Not Fail"

Telephone (code 503) 686-3555
Eugene, Oregon 97403

November 16, 1977

Rosemary Wilson
Office of Education
Department of Health, Education
and Welfare
7th and D St. SW
Washington, D.C. 20202

Dear Ms. Wilson:

The critical Follow Through issue is a moral one. We have demonstrated the capacity to teach you and other educators about teaching "poor kids", turning them on, and assuring that they catch up to their middle-class peers in academic skills. Our Follow Through achievements, however, don't show what we are actually capable of doing, because we do not have fully-implemented sites - only moderately implemented sites. The Follow Through guidelines have never permitted the kind of total system support needed to provide a full-fledged demonstration of what poor kids can achieve in grades K-3 if they receive a fully implemented, uniform, Direct Instruction approach (with trained teachers, supervisors, and directors).

The moral issue centers on this question: What does Follow Through stand for? Is it simply an experiment on human beings which has no concern for what the experiment might reveal for the millions of other "poor kids" who have serious educational needs? Is your office even remotely justified in treating all sponsors as equals and shifting (in the past two years) the emphasis from sponsors to individual sites? Or is this move designed to detract from the issues of effective approaches and make it seem that "every approach is capable or producing good results"? We would not have engaged in Follow Through for the past ten years if we had thought that there would be no attempt to weed out the inadequate model approaches, to educate both the public and the educational community about how to be effective, and to disseminate information on how to be an effective model? Furthermore, we would not have perpetuated our model if it proved to be a hoax; rather, we would have quit Follow Through with abject apology if the results had shown that our approach did not work any better than that of the typical Title I program and produced kids who performed only at the 18th percentile in reading and arithmetic by the end of the third grade.

According to your letter of October 28, 1977, "No single instructional approach, including that of Direct Instruction, was found to be consistently effective in all of the projects where it was tried and evaluated. Therefore, no overall claim of its effectiveness can be supported."

Rosemary Wilson
 November 16, 1977
 page 2

This statement flies in the face of the USOE's own Evaluation Synthesis (John Evans' office) indicating that there is one model generally effective in basic skills, cognitive skills, and on affective measures - Direct Instruction. Your statement also flies in the face of the Abt Report data - not merely in the face of data showing the relative superiority of Direct Instruction over comparison groups, but also in the face of absolute grade norm data that shows Direct Instruction to be the only approach to bring third graders at or near grade level in reading, arithmetic, spelling, and language.

Your conclusion that no single instructional approach was found to be consistently effective is spurious because: (1) Grand Rapids is included as a Direct Instruction site (for two data points); and (2) the conclusion confounds program effect variability with control group variability. Variability in program implementation must take into account absolute grade-level performance. Grade-level performance is clearly as relevant as statistical significance because if all kids were performing at grade level (on the average), there would have been no need for Follow Through. In fact, much of the rhetoric that led to Follow Through by Robert Kennedy, Commissioner Howe, Bob Egbert and Jack Hughes, focused on the fact that poor kids perform well below grade level and are therefore preempted from "higher education" and its concomitants.

The so-called "variability" among the Direct Instruction sites is accounted for first by the unwarranted inclusion of Grand Rapids as a Direct Instruction site for two cohorts. Grand Rapids is not a Direct Instruction site. It is a Lola Davis site. You know that for most of 1972 and 1973, we did not function as a sponsor for Grand Rapids, that we did not receive funds for servicing Grand Rapids after the spring of 1973 when relationships with Grand Rapids were severed. You also know that Grand Rapids was not functioning as an implemented Direct Instruction site before the 1972-73 school year. It was not implemented - not because of our lack of effort - but because the director had different ideas about what and how to teach. You were involved in this situation and know quite well our position and the history. Grand Rapids is the only Direct Instruction K-starting site that performs relatively low in absolute performance. In fact, Grand Rapids performs an average of 1/2 standard deviation below the mean of other Direct Instruction K-starting sites. Examine the following summary tables that are based on data from Abt 3 and Abt 4, which show medium grade norms for Direct Instruction sites.

	<u>Total Reading</u>	<u>Total Math</u>	<u>Spelling</u>	<u>Language</u>
	3.7	4.2	4.3	4.8
Norm(3.5) →	3.5	4.0	4.3	4.8
	3.4	4.0	3.9	4.7
	3.4	3.9	3.9	4.6
	3.3	3.9	3.7	4.5
	3.2	3.8	Norm → 3.6	4.2
	3.1	3.8	3.4	4.1
	3.1	Norm → 3.75	3.4	Norm → 4.0
	(2.8)	3.7	3.4	(2.8)
	(2.8)	(3.0)	(3.0)	(2.6)
		(2.8)	(3.0)	

() Grand Rapids Norm = 50th percentile

Rosemary Wilson
November 16, 1977
page 3

Although there is a variability, note that most of the variability is above the national norm (with the removal of Grand Rapids). Grand Rapids is the only site that consistently falls below national norm. Even the most casual inspection of these data suggests that Grand Rapids is "different". And you know that it is different because it has not been implemented as a Direct Instruction site.

In absolute data, Direct Instruction, even with Grand Rapids included, out-performed the other sponsors by one-fourth to one full standard deviation on MAT Total Reading, Language, Total Math, and Spelling. Furthermore, we can document the fact that what can be achieved through the Direct Instruction approach is only partly reflected in the table above. We can show, for example, that implementation was not fully achieved in any site and that we can do better. Our questions to you are: Wouldn't it behoove Follow Through to pursue the possibility that what we are saying is true? Wouldn't it be of potential value for educating disadvantaged kids to know what really could be done with optimal implementation? Wouldn't it be valuable to have a bench mark, a standard of excellence that establishes what can be done and that can therefore serve as a goal for other projects and schools more generally?

In your letter, you assert: "Since 1968, the program emphasis ... has been on the cooperative development and evaluation of Follow Through projects, not models ..."

There are two problems with this statement. The first is that it is historically false. The second is that it places you on the horns of a serious dilemma.

1. The historical facts are these:

In his history of Follow Through prepared for the Brookings Conference on Planned Variation, Egbert asserted while referring to the planning going on in 1967: "With such limited funds, it seemed sensible to change Follow Through's primary purpose from 'service to children' to 'finding out what works'. ... Follow Through now focused its attention on developing, examining, and refining alternative approaches to the education and development of young disadvantaged children."

Fairley's five-year plan in 1971 was aimed at disseminating successful models into Title I applications, a project that was terminated because the data were not yet in. The following quote is taken from the SRI Administrative History (p. F28, quoted from Haney, 1977, p. 36): "It was clear from Mr. Fairley's comments that his interest was largely upon overall program effects and an identification of 'best' individual sponsors to guide decisions about future Follow Through program scope." [Emphasis added.]

The following statement is found in the first page of the Abt IV Report on the evaluation of Follow Through which has been accepted by the Office of Education and due for release to Congress shortly: "As part of this evaluation, the U.S. Office of Education commissioned Abt Associates in 1972 to

Rosemary Wilson
November 16, 1977
page 4

analyze the data generated by the extensive program of testing and interviewing which was part of Follow Through and to draw from them appropriate conclusions about the effectiveness of the various models' approaches to compensatory education." (Abt IV, 1977, p. xxiii) [Emphasis added.]

In his history of Follow Through in Rivlin and Timpane's book on Planned Variation in Education, Elmore states: "The idea of planned variation experimentation adopted by Follow Through was, of course, very different. Instead of specifying dimensions of variation and attempting to limit the experiment to those most susceptible to definition and replication, Follow Through administrators tried to select promising, innovative program models, leaving the question of how these models differed to a later time." (Rivlin & Timpane, p. 36)

Garry McDaniels, former director of Follow Through research, makes these statements discussing the goals of the Follow Through evaluation:

"Do the various educational strategies used in Follow Through have different effects, and do the effects endure?" (Rivlin and Timpane, p. 47) ... "The planned variation between models of various sponsors is considered to be a more reliable means of assessment than natural variation..." (Rivlin and Timpane, p. 48)

You assert that the emphasis of support is and has been on individual sites, not sponsors. This position, however, does not explain why Follow Through is organized as it is. If the emphasis is not on sponsor, why isn't there a "random variation" of approaches used in different sites? Why were sponsors introduced in the first place? The notion of planned variation makes very little sense and becomes self-contradictory if in fact there was no attempt to draw conclusions from a controlled pattern of implementation across sites. Furthermore, the National Office has commissioned studies to document the fact that different sponsors produce different patterns of implementation:

The Nero implementation study (1975) shows that sponsors are discriminable, that classroom features clearly identify different sponsors, and that there is a different pattern of performance across different sponsors. (What was the purpose of this investigation if there was no attempt to find out about sponsors, and their implementation across sites.)

Similarly, the Stallings and Katkowitz study, funded by USOE, showed through classroom observations that sponsors were discriminable in their implementation at different sites and at different grade levels. The study supported the conclusion that sponsorship is a viable concept and that sponsor implementation does lead to observable differences in classroom outcomes and in performance outcomes.

Rosemary Wilson
November 16, 1977
page 5

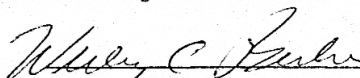
2. One horn of your dilemma is this:

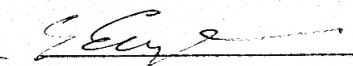
If you deny (which you apparently have done) the sponsor concept of Follow Through and deny that what was to be measured was sponsor capability, not just the capacity of different sites to produce programs, you make it transparently clear that Follow Through actually conned individual sponsors into thinking that their efforts would be reinforced if they produced positive results and that good approaches - those with uniform potential - would be disseminated. You admit that Follow Through is a sham and that all sponsors would be equals - those achieving student performance at the 50th percentile and those at the 15th percentile.

On the other hand, if you accept the sponsorship philosophy of Follow Through and the idea that the game was, from the beginning, an attempt to find out about methods, programs, approaches that are effective across different sites, different types of kids, and different ethnic groups, you are faced with the conclusion that you are now supporting many sponsors that have shown precisely no capacity to produce, and in fact are showing alarmingly consistent negative results.

We do not envy your position. However, we do not have to reinforce it and continue to be a part of an obvious travesty. We wanted to show what could be done for the kids. After nearly ten years, we find that although we succeeded, we have been rejected - not merely by the outside educational establishment from whom rejection would be a natural response - but from the agency that has funded us, that required us to hold to a moratorium on publishing comparative data before 1975, that repeatedly suggested possible expanded funding of successful models, that posed as something more elegant than a fancy Title I program. It seems apparent, however, that Follow Through at the National Level, has become a bureaucracy with no apparent advocacy for the needs of children.

What other form of significant protest do we have other than quitting Follow Through, other than by severing association with the kind of non-relevant educational agencies that we have tried to fight during the past years?


Wesley C. Becker
Professor of Special Education


Siegfried Engelmann
Professor of Special Education

cc: Richard Fairley
Thomas Minter
Ernest Boyer
Robert Egbert
Mary Berry
Representative Weaver
Representative Ullmann
Senator Packwood
Senator Hatfield
Dean R. Gilberts
Senator Nelson

Rosemary Wilson
November 16, 1977
page 6

References

Abt Associates, Education as Experimentation: A Planned Variation Model, Vol. 3, 1976, Vol. 4, 1977, Cambridge, Mass. Abt Associates.

Haney, W. The Follow Through Planned Variation Experiment, Vol. 5: A Technical History of the National Follow Through Evaluation, Cambridge Mass., Huron Institute, 1977.

Nero and Associates, A Description of Follow Through Implementation Processes, Portland, OR, Nero and Associates, 1975.

Egbert, R., "Planned Variation in Follow Through," unpublished manuscript, presented to Brookings Institution's Panel on Social Experimentation. Washington, D.C.; April 1973.



DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

DEC 19 1977

Dr. Wesley C. Becker
Dr. Siegfried Engelmann
Professors of Special Education
University of Oregon
College of Education
Eugene, Oregon 97403

Dear Sirs:

This is in response to your letter to me of November 16, 1977, in which you state among voluminous other assertions, that a statement I had made in an earlier reply to you of October 28, 1977, was "historically false." Although, as Director of the Follow Through program, I have many constructive demands upon my time that prevent my entering into what now threatens to become an endless exchange of unproductive, widely distributed, correspondence between you and myself; I am compelled to respond to that charge of falsehood.

Since, on October 28, I was writing a letter to you and not a position paper as you have now done, I felt it unnecessary to provide annotation of the sources upon which I based my statements. I do so now, however, in view of this charge which you have so widely circulated.

The Follow Through Program Manual (draft), dated February 24, 1969, was prepared and used while this program was under the direction of Dr. Robert Egbert, whose later unpublished writings your paper quotes extensively. The program manual contains on page 2, the following statement under B. Planned Variation in a Context of Comprehensive Services.

"In February of 1968, the U.S. Office of Education invited a limited number of communities (recommended by State officials) to participate in a cooperative enterprise to develop and evaluate comprehensive Follow Through projects, each of which incorporates one of the alternative 'program approaches' as part of its comprehensive Follow Through project. Generally, each of the current program sponsors concentrates on only a portion of the total Follow Through project. The remainder of the program is developed by the local community with consultant assistance."

Page 2 - Dr. Wesley C. Becker and Dr. Siegfried Engelmann

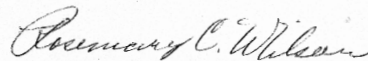
Based on the above quotation, I made the following statement in response to your suggestion that this office recommend your Direct Instruction Model to the Joint Dissemination Review Panel.

"This is not reasonable in that it directly conflicts with OE's decision that projects, not "models" would be presented to the JDRP. This was a carefully considered decision based on the following facts.

- a. Since 1968 the program emphasis has been on the cooperative development and evaluation of comprehensive Follow Through projects not models.
- b. No single instructional approach, including that of Direct Instruction, was found to be consistently effective in all of the projects where it was tried and evaluated. Therefore, no overall claim of its effectiveness can be supported."

In 1968, I was not in the employ of the U.S. Office of Education. The draft program manual to which I refer, however, contained the guidelines used by the Follow Through program and its grantees to govern their operations until June 21, 1974 when, under my direction, the Follow Through program published in the Federal Register as an interim final regulation, its first official regulation (39 FR 22342 et. seq.)

I add that "quitting Follow Through" as a form of "significant protest" was your decision. In my opinion, it is a very destructive one to the Follow Through program and to the "poor kids" to whom you profess commitment.


Rosemary C. Wilson

Obviously, Rosemary Wilson lied, issuing the same fabrication that House and Glass asserted about Follow Through being designed to assess the aggregate performance of the models, not to compare individual sponsors. Equally obvious, Wilson was told by her superiors what lies she would present about the intent of the Follow Through evaluation. The only real question was how high in the bureaucracy the deception went.

The answer came in 1978, in the form of a letter from U.S. Commissioner of Education, Ernest Boyer, to Senator Bob Packwood from Oregon, one of the politicians Wes had contacted. This letter provides no doubt about Rosemary Wilson lying. It also delivers the twisted justification for rejecting the focus on sponsors. The letter leaves no doubt that the decision came from the top of the food chain, the Commissioner of Education. This letter is frank and honest, but contains desperately confused arguments.

Following is the letter from Packwood to Wes and the accompanying letter from Boyer to Packwood.

RUSSELL B. LONG, LA., CHAIRMAN

HERMAN E. TALMADGE, GA.	CARL T. CURTIS, NEBR.
SEN. SAM BRIDGEMAN, CONN.	CLIFFORD P. HANSEN, WYO.
HARRY F. BYRD, JR., VA.	ROBERT J. DOLE, KANS.
GAYLORD NELSON, WIS.	BOB PACKWOOD, OREG.
MIKE GRAVEL, ALASKA	WILLIAM V. ROTH, JR., DEL.
LLOYD BENTSEN, TEX.	PAUL LAXALT, NEV.
WILLIAM D. HATHAWAY, MAINE	JOHN C. DANFORTH, MO.
FLOYD K. HASKELL, COLO.	
SPARK M. MATSUNAGA, HAWAII	
DANIEL PATRICK MOYNIHAN, N.Y.	

MICHAEL STERN, STAFF DIRECTOR
GORDON S. GILMAN, CHIEF MINORITY COUNSEL

United States Senate

COMMITTEE ON FINANCE
WASHINGTON, D.C. 20510

April 19, 1978

Wesley C. Becker, Ph.D.
University of Oregon
College of Education
Department of Special Education
Eugene, Oregon 97403

Dear Dr. Becker:

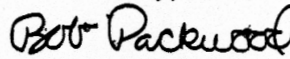
At last I received a response from the Office of Education regarding my questions on the Follow Through Program. I have enclosed their response to me with this letter. It would appear from the letter that the Office of Education and the President feel that (1) the Follow Through Program was generally not successful in producing any definitive results; and (2) that they are therefore abandoning any continuation of the program.

I would certainly welcome your comments on the points that they make in their letter.

As you know, the Subcommittee on Employment, Poverty and Migratory Labor of the Senate Human Resources Committee held hearings on S.2090 which proposed to extend the Head Start and Follow Through Programs for another three years. The Subcommittee will be marking up the bill next week.

I would greatly appreciate receiving any thoughts you have on this bill, and how I might be of any assistance to you in your program efforts. Thank you for contacting me, and I look forward to reading any comments you may have.

Cordially,



BOB PACKWOOD

BP/gzp



DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
OFFICE OF EDUCATION
WASHINGTON, D.C. 20202

MAR 31 1978

Honorable Bob Packwood
United States Senate
Washington, D.C.

Dear Senator Packwood:

Thank you for your letter inquiring about the Follow Through Program. I will speak to your specific questions in order:

1. Since the beginning of Follow Through in 1968, the central emphasis has been on models. A principal purpose of the program has been to identify and develop alternative models or approaches to compensatory education and assess their relative effectiveness through a major evaluation study which compared the performance of Follow Through children with comparable children in non-Follow Through projects over a period of several years. That study had just been completed. A summary of part of the final result, which was sent to the Congress last summer, is enclosed; a more comprehensive final report will be issued in the next few weeks. The evaluation found that only one of the 22 models which were assessed in the evaluation consistently produced positive outcomes. The central finding of the evaluation was that there was substantial variation in effectiveness among the sites in almost all of the models. Accordingly (and this speaks in part to your second question), we are funding 21 of the successful sites as demonstration sites this year so that other schools and educators will learn about, understand, and hopefully adopt the successful activities and procedures taking place in these effective sites. In summary, while the initial emphasis of the program was on designing and implementing models, the evaluation results forced us to shift attention more to successful individual projects. If the evaluation findings had indicated that the various models tested were either generally effective or generally ineffective, then the subsequent demonstration and dissemination activities could have proceeded along model rather than individual project lines. However, with the exception of the Becker-Engelmann Direct Instruction model, this was not the case.

2. The President's budget recommendation for FY 1979 is to begin a phase-out of the Follow Through program. The experiment to assess this particular set of compensatory education models has been completed. It

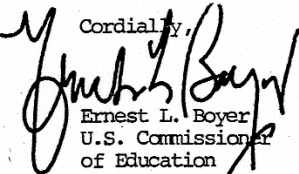
Page 2 - Honorable Bob Packwood

is felt that future research and development efforts of this nature ought to be carried out by the National Institute of Education. The Administration is proposing an increase of \$600 million in Title I of the Elementary and Secondary Education Act which will substantially expand the kind of direct compensatory education assistance which children in the Follow Through sites are receiving.

With respect to your question about funding selected models on an expanded scale, with the exception of the Direct Instruction Model this could not responsibly be done since, as I noted above, positive evaluation evidence for the other models is lacking. The same problem applies to your question about funding sponsors to disseminate information on Follow Through programs on the college and in-service levels. Since only one of the models, and therefore only one of the sponsors, 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 which carried the implication that such models could be expected to produce generally positive outcomes.

I hope this information answers your questions. If I can be of any further assistance, to you, please let me know.

Cordially,



Ernest L. Boyer
U.S. Commissioner
of Education

The first sentence of point 1 in Boyer's letter contradicts the assertion by Wilson, House, and Glass about whether Follow Through was designed to find successful models or to evaluate the aggregate of models. "*Since the beginning of Follow Through in 1968, the central emphasis has been on models.*"

Boyer freely admits that policy makers accepted the data as valid. Several references in his letter indicate that he had no doubt that only one model was highly successful, which means that he was aware of facts that had never been shared with states and school districts.

The ultimate conclusion Boyer drew was that if there was only one successful model, it should be treated like all the other models. In response to the question about funding *selected* models, Boyer's logic seems to be that somehow such funding would be irresponsible because there were not selected models, *only one selected model*. So rather than fund that model, the Office of Education assumed it was equitable to treat all models the same and simply promote selected sites. Imagine spending half a billion dollars to draw this conclusion.

The effect Boyer presumed would happen is naïve: "*... we are funding 21 of the successful sites as demonstration sites this year so that other schools and educators will learn about, understand, and hopefully adopt the successful activities and procedures taking place in these effective sites.*"

Boyer had data that the effective non-DI schools were aberrations and that they were so elusive that the sponsors could not even train their other schools to do what the successful school did. If there was any validity to the notion that people would visit a dissemination model for High Scope and be able to implement as well as the school visited, the sponsor would have been the first to know about this excellent site and therefore the first to try to disseminate in his other sites. This dissemination failed. The successful school remained an outlier. Therefore, there would be no hope of visiting schools being able to replicate the procedures of this school. In fact, the National Diffusion Network (NDN) did not create more than a handful of success stories for failed schools.

Schools from High Scope and other failed models were disseminated for one reason: to preserve at least a modicum of credibility to all the favored ideas and practices of mainline educational thought. If everybody failed, at least Stallings, Piaget, and the rationale that drove at least 19 of 22 models would not be shown to be grossly inferior to the ideas and practices that innervated DI.

In terms of morality, Boyer's decision not to permit sponsors to disseminate was brutal. Why wouldn't it have been possible to fund us as a model and fund sites from other models? The consistent performance of our model affirmed that our techniques and programs were replicable and that with

proper training teachers in failed schools could succeed. Why wouldn't that information be important enough to disseminate? Why did the government feel that it had to initiate some form of affirmative action to keep failed models floating?

Boyer admits that the results didn't come out the way experts predicted. Policy makers didn't have the vision of only one program excelling in basic skills and cognitive skills, or the same program excelling in reading, spelling, and math. They were not prepared for the possibility that this program would also have children with the strongest self-image.

MANIPULATED DATA?

Earlier, I suggested the possibility that policy makers tried to sour our data by purposely including Grand Rapids as one of our sites. Two sentences in Boyer's letter may confirm this suspicion:

The evaluation found that only one of the 22 models which were assessed in the evaluation consistently produced positive outcomes. The central finding of the evaluation was that there was substantial variation in effectiveness among the sites in almost all of the models.

If these sentences are considered literally, they imply that in the original report Boyer received, not all of the models had variation. There was substantial variation in *almost all of the models*; however, one *consistently produced positive outcomes*. Possibly the addition of Grand Rapids was an intentional manipulation to create variation and thereby make it possible for conspirators Glass and House to present their argument on variability within models.

Whether or not the data were manipulated, there had been a fairly extensive plot to assure that various bureaucrats told a consistent story about the intent of the Follow Through evaluation and did not contradict one another (at least until Boyer blew it with his letter to Packwood). The parties included House, Glass, the Ford Foundation, Rosemary Wilson, Follow Through, the National Institute of Education, and the Office of Education, all the way to the top.

The extent to which the distorted account of Follow Through prevailed over truth was partly revealed by an online outline of significant educational events that occurred during the 1960s and '70s.

I discovered the outline while doing research for this chapter. The outline was for a college course at Illinois State University, Political Science

233: Politics and Public Policy. The instructor was Gary Klass. The outline went into some detail about the Coleman Report and the Pettigrew interpretation of the Report, which led to bussing. The outline covered the failure of bussing and the failure of Head Start. It did have a note that a preschool produced benefits. That preschool was the Perry Preschool, which is High Scope.

Following the endorsement of High Scope was a heading, *Other Studies*, followed by a one-line reference to Follow Through:

Compensatory education programs show no effect.

Done.

If people like Klass didn't have a clue, the campaign to bury the truth about Follow Through had to be pretty effective.

Another way of measuring the effectiveness of the historical distortions of Follow Through is to tell the truth. On three occasions I talked about our model to non-educational audiences. One was a Chamber of Commerce; the others were business groups that supported different school efforts. The responses were the same. After I gave the facts, at least one member of the audience would say something to the effect, "You're telling us that you achieved all these things in Follow Through but professionals in the field rejected your model. I know some people in education, and they are well informed and committed to do a good job. But you're saying that they would purposely ignore actual facts about student achievement. I've always believed that if someone builds a better mousetrap it will sell. You're telling us that's not true in education. I find that hard to believe. I also question whether the educational system would plot against your program if it was as successful as you claim it is."

After the third talk I resolved never to do it again, and I haven't. But I've had the same experience dealing with administrators—the frown, the head-shake, and the confession, "I find that hard to believe."

The saddest part of the Office of Education's conspiracy to propagate lies and intellectual casuistry is that it makes a mockery of the vision that Robert Kennedy had when he argued for evaluation—so that educators would make sensible responses based on the outcome data.

That could have happened only if the Follow Through data were properly disseminated; however, such dissemination was unpalatable to those who had power. Stated differently, on the balance scale of reality, the weight of Jackie, Alan, and all the other poverty kids on one side didn't come close to balancing the weight of influential people, their prejudices, and their financial interests on the other.

However, this failed system could have benefited in the long run if it had

an understanding that the process of creating effective programs is greatly different from the approach they used; that there was a theory that explained the details; and that there were people who would have been glad to share whatever they knew about efficient ways of doing it.

During the tumult of 1977–79, I did not participate in the political side of things beyond the letter to Rosemary Wilson. I continued to develop instructional programs and work with our remaining sites, as Wes took on the bulls and the bears. Both left scars on him. After Packwood sent him the Boyer letter, Wes wrote Packwood again. Here's the last part of that letter.

The basic problem we face is that the most popular models in education today (those based on open classrooms, Piagetian ideas, language experience, and individualized instruction) failed in Follow Through. As a result there are many forces in the educational establishment seeking to hide the fact that Direct Instruction, developed by a guy who doesn't even have a doctorate or a degree in education, actually did the job. To keep those promoting popular approaches from hiding very important outcomes to save their own preconceptions will take formidable help from persons like yourself. We hope it is not too late.

Sorry, Wes. It was too late. The truth about Follow Through was silently drowned like an unwanted kitten, and nobody protested. Outfits like the NAACP and other advocacy or community-action groups should have been outraged, but they were conspicuously mute, apparently lacking the means, knowledge, or commitment needed to be more than paper advocates. The drowning was a complete success.