PET and the Pendulum: Faddism in Education and How to Stop It
Author(s): Robert E. Slavin
Source: *The Phi Delta Kappan*, Vol. 70, No. 10 (Jun., 1989), pp. 752-758
Published by: Phi Delta Kappa International
Stable URL: http://www.jstor.org/stable/20404026

PET and the Pendulum: Faddism in Education and How to Stop It

If education is ever to make serious generational progress, educators must somehow stop the pendulum by focusing their efforts to improve education on programs that are effective, rather than on those that are merely new and sound good, Mr. Slavin asserts.

BY ROBERT E. SLAVIN

EDUCATIONAL innovation is famous for its cycle of early enthusiasm, widespread dissemination, subsequent disappointment, and eventual decline — the classic swing of the pendulum. Of course, a similar pattern exists in most applied fields, but many nonetheless exhibit steady generational progress that is far more important than the latest fad. For example, there are fads in medicine, agriculture, and engineering, but they occur against a backdrop of steady, widely acknowledged, and irreversible progress.

Generational progress does occur in education, but it is usually a product of changes in society, rather than changes in educational techniques themselves. For example, the clearly beneficial trend toward desegregation and more equal treatment of minorities represents true generational progress, but it arose from social and legal changes, not from educational innovation. More often, education resembles such fields as fashion and design, in which change mirrors shifts in taste and social climate and is not usually thought of as true progress.

ROBERT E. SLAVIN is director of the Elementary School Program at the Center for Research on Elementary and Middle Schools, Johns Hopkins University, Baltimore. *He wishes to thank Gary Gottfredson, Nancy Madden, Garrett Mandeville, and Janelle Rivers for their comments on a draft of this*

One major factor inhibiting systematic progress in education is the lack of agreement about what constitutes progress and what constitutes adequate evidence to support action. Other fields have clear indicators of progress – such as yield per acre, mortality per 1,000 patients, or return on investment – and they have clear roles for research and development. When the Food and Drug Administration (FDA) approves a drug, physicians confidently prescribe it. In education, there is no FDA.

If education is ever to stop the swinging of the pendulum and make significant progress in increasing student achievement, it must first change the ground rules under which innovations are selected, implemented, evaluated, and institutionalized. In this article I discuss the dynamics of the educational pendulum, using as a case in point the Madeline Hunter phenomenon that is currently cresting in American schools. I propose a series of steps that school districts and government and other funding agencies could take to provide an infrastructure capable of promoting lasting and beneficial change in educational practice.

THE PENDULUM IN PRACTICE

One of the most important reasons for the continuing existence of the educational pendulum is that educators rarely wait for or demand hard evidence before adopting new practices on a wide scale. Of course, every innovator claims research support for his or her methods; at a minimum, there is usually a "gee whiz" story or two about a school or district that was "turned around" by the innovation. Alternatively, a developer may claim that, while the program itself has not been formally evaluated, the principles on which it is based are supported by research.

This latter claim has been made with considerable justification by Hunter, perhaps the most popular educational trainer of our time.[1] Hunter's model, often called Instructional Theory into Practice (ITIP), describes critical elements of lessons (anticipatory set, statement of objectives, lesson presentation with frequent checks for understanding, guided practice, independent practice, and closure) and also emphasizes such concepts as activating prior knowledge, teaching for transfer, and using cognitive strategies.

The Hunter model is appealing, practical, and well-grounded in educational and psychological theory. But does it work? Fortunately, there is no ambiguity about Hunter's objective, which is stated in the title of one of her books, *Teach More – Faster!*[2] Asking whether Hunter's model works clearly means asking, "Do students taught by ITIP-trained teachers learn more (and learn faster) than students taught by other teachers?"

Some innovations (e.g., the open school) are difficult to evaluate because their developers disdain traditional measures of achievement. In the case of ITIP, however, it would be hard to make an argument that "teaching more and faster" would fail to affect scores on standardized or criterion-referenced tests. Since almost all schools administer such tests and since the number of teachers trained in ITIP methods must be in the hundreds of thousands, one might imagine that there would be dozens of studies comparing ITIP to control methods.

Amazingly, such is not the case. The first published article that evaluated ITIP was a study by Jane Stallings,[3] which appeared 16 years after publication of the 1969 edition of *Teach More – Faster!* Stallings studied a single school in Napa, California, for three years and found small achievement gains in the first two years that disappeared in the third year.[4] Even those early gains were called into question by the fact that a control group showed a similar pattern of gains.[5]

Although clearly disappointing, the Napa study could not provide conclusive evidence that ITIP was *not* effective. There were some gains in the early years, the control group was not truly equivalent to the experimental group, and there was some possibility that teachers of the control group heard about and implemented parts of the ITIP model.[6] Hunter criticized the Napa study on the grounds that close supervision of the implementation of her model by the teachers did not continue into the third year, that the program had been implemented too rigidly, and that the quality of training was not high.[7]

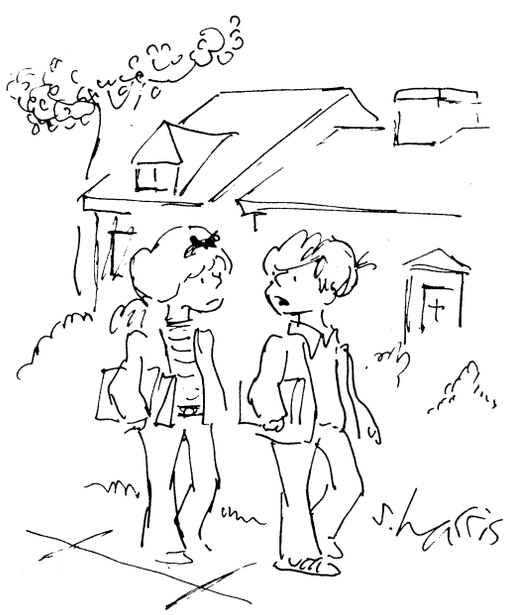
THE PET IN SOUTH CAROLINA

Nineteen years after the Hunter phenomenon began in earnest, a high-quality, large-scale evaluation of ITIP has finally appeared. At the annual meeting of the American Educational Research Association (AERA) in New Orleans in 1988, a symposium was held on the statewide implementation of ITIP in South Carolina, which included an evaluation of the achievement effects of the program over a three-year period. The South Carolina model, called the Program for Effective Teaching (PET), is an ambitious undertaking intended to train teachers in Hunter's methods in 87 of the state's 91 school districts. John Tudor of the South Carolina Department of Education told the AERA how 15,362 teachers (half of the teachers in the state), 2,033 administrators, and 1,118 trainers had been trained. He also described how the state provided extended training, offered in-class follow-up training, and altered administrative structures to insure high-quality implementation.[8]

Nineteen years after the Hunter phenomenon began in earnest, a high-quality, large-scale evaluation of ITIP has appeared.

Two PET trainers then described in more detail the impressive steps taken in their district to insure the success of the training, including involving administrators and providing follow-up services, peer coaching, and coaching from trainers.[9] A researcher from the University of South Carolina presented data indicating that teachers trained in the Hunter model were overwhelmingly positive toward it.[10]

Then came the bad news. Garrett Mandeville, another University of South Carolina researcher, found no important differences in achievement between students of PET-trained teachers and students of other teachers, after controlling for prior achievement and for socioeconomic status (defined by eligibility for free or reduced-cost lunches). In fact, in classes in which teachers had been trained



"The Golden Age of Hollywood, the Golden Age of TV, the Golden Age of Rock . . . we missed them all."

in PET two to three years before the posttesting was conducted, students scored slightly *worse* than did students of untrained teachers (by about 9% of an adjusted standard deviation). Further analyses revealed that achievement scores among students of PET-trained teachers were not significantly related to the amount or perceived quality of coaching, attitude toward PET, self-reported use of PET concepts and lesson plans, motivation for training, or other variables. Mandeville described the painstaking procedures used to insure that the students of trained and untrained teachers were identical in every important way and that trained and untrained teachers were similar in experience, age, class size, number of repeaters, and other variables.[11]

Jane Stallings and I critiqued the PET symposium papers at the AERA meeting, focusing on the high quality of the state plan for disseminating Hunter's methods, the high quality of the training, the extraordinary quality of the evaluation, and the disappointing lack of effects. However, in response to a question from the audience, one of the PET trainers described her excitement about the state's plan to train the remaining 15,000 teachers in South Carolina. In fact, the state is proceeding with its training plans, essentially ignoring the results of its own evaluation. If the study had any impact on the state's plans, it was to focus attention on the quality of coaching, not on the notion of whether or not to go ahead with the program.[12]

The South Carolina study is not the only large-scale evaluation of Hunter's ITIP. James Donovan, David Sousa, and Herbert Walberg reported an evaluation of the program in grades 3, 6, 9, and 11 involving 35 trained and 29 untrained teachers in West Orange, New Jersey. As in the South Carolina and Napa studies, adjusted achievement scores for students of trained and untrained teachers were nearly identical.[13]

Taken together, the results of the Napa, West Orange, and South Carolina evaluations of ITIP offer little hope that the Hunter approach will produce any improvement in student achievement. Certainly this conclusion is important in its own right, but the questions it raises are even more important: Why did it take 19 years for an adequate evaluation of ITIP to be conducted? Why did American education jump so enthusiastically onto the Hunter bandwagon in the absence of evaluation data? Most important, how can we avoid continually repeating this pattern of educational change?

The case of ITIP provides a vivid example of the educational pendulum in action. It is particularly instructive because of the breadth of adoption of the ITIP model, the lack of evaluative evidence until very recently, and the unambiguous failure of the program to affect the only variable it is designed to affect: student achievement.

If change in education is ever to produce progress, rather than just another swing of the pendulum, we need to understand why and how the pendulum operates. The process can be considered in two phases: the upswing and the downswing.

THE UPSWING

1. *Program is proposed.* Usually the pendulum begins its upswing with the publication of the idea in a popular education periodical, such as *Instructor*, *Educational Leadership*, or *Learning*, or the publication of a popular book. In the case of ITIP, the publication of a series of small books between 1967 and 1969 and of articles in *Instructor* gave the pendulum its initial impetus.

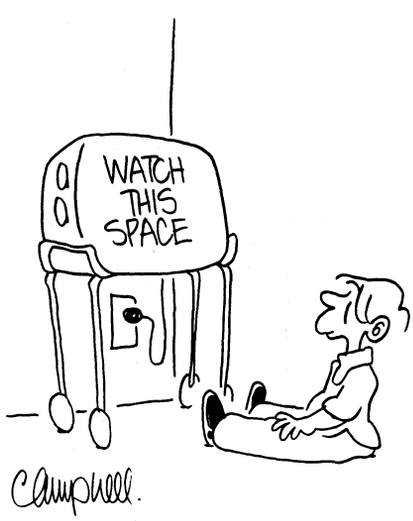
2. *Program is piloted.* "Preliminary evidence" of effectiveness often takes the form of "gee whiz" reports. Early in the dissemination process, promising results of studies may be reported from one or more school districts. These studies are almost always badly flawed. At best, the data represent improvements over the previous year's scores; at worst, they may be completely anecdotal. In the case of ITIP, an unpublished, unavailable early evaluation in Long Beach, California, served the "gee whiz" function.

3. *Program is introduced in innovative districts.* Certain school districts have a reputation for trying out the latest innovations. There is an often-told joke about such districts: "If the state superintendent said, 'Go to hell,' District X would make sure it got there first."

4. *Program becomes "hot topic" among staff developers.* The substantial corps of staff developers, curriculum supervisors, and others whose job is to introduce the latest innovations to their school districts are always willing to give the pendulum a push on the upswing. These professionals are usually expected to know about all the latest developments; however, they are rarely rewarded for sticking with a particular innovation for many years until it is well-established and well-evaluated within a district. As a result, this year's hot instructional methods eventually become next year's old news. When a method becomes the talk of the convention in such organizations as the Association for Supervision and Curriculum Development, the American Association of School Administrators, the International Reading Association, or the National Council of Teachers of Mathematics, it has arrived at this critical stage.

> Why did American education jump so enthusiastically onto the Hunter bandwagon in the absence of evaluation data?

5. *Program expands rapidly.* As soon as a program becomes "this year's thing," training expands rapidly. Most of the training consists of one-day (or shorter) inservice workshops given to large audiences of teachers by the developers of the program or by staff development specialists who just a year earlier were equally enthusiastic about something completely different. Classroom follow-



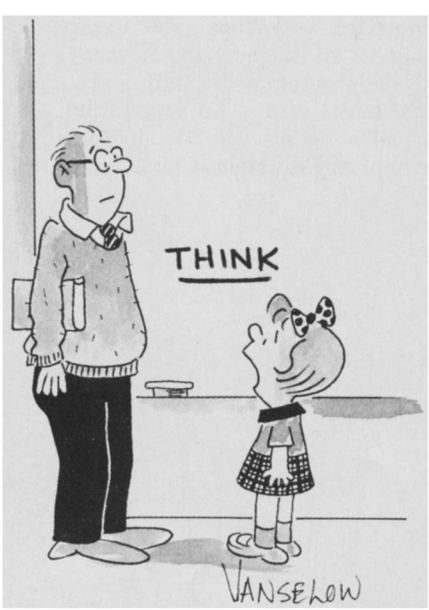
up to insure that the program is being properly implemented is rare at this stage in the pendulum's swing.

6. *Controlled evaluations begin*. Serious, controlled evaluations of the program usually do not begin until the program is already expanding rapidly. There are several reasons for the delay. First, it is hard to evaluate a program properly until a lot of teachers are using it. Second, there is a lag between the initial awareness that a program is becoming widely used and the appearance of an opportunity to study it. This lag is sometimes caused by the need to find grants to support an evaluation. For example, in 1980 those in charge of the federal Follow Through program decided to implement and evaluate promising programs directed at increasing time-on-task and implementing the findings of research on teacher effectiveness. This was the program that funded the Napa study, for which the final report wasn't published until 1987.

Ironically, controlled evaluations often begin at the peak of a program's popularity, the very top of the upswing of the pendulum. Thus the results of these evaluations may appear when the innovation is on the wane for other reasons.

THE DOWNSWING

7. *Innovative districts move on to other programs*. The downswing of the pendulum begins when the innovative districts that first tried the new method find something newer. For example, many of the early ITIP enthusiasts are now training to use innovative programs in cooperative learning and the teaching of thinking skills, which (at this writing) are on the upswing. Moving on to new methods is not, of course, just an indication of faddism; it may reflect concerns about real or perceived problems with the method.



"We did that last year – how come we have to do it again this year?"

> Until a program becomes widely known and used, critical articles are of little interest and would not be published.

8. *Complaints surface in professional publications*. Every innovation, no matter how effective or popular, has its detractors. Until a program becomes widely known and used, however, critical articles are of little interest and would not be published even if they were written. After all, who cares if someone dislikes a method that few people are using or discussing?

9. *Preliminary evaluations are disappointing*. At about this point in the downswing, early results of evaluations begin to appear in the literature or become known in some other way. Often, these results are disappointing. However, the early studies may be flawed and therefore can be dismissed. Such was the case with ITIP. The disappointing findings of the Napa study could indeed be criticized because the study was small, the control group was not initially equivalent to the experimental group, and so on. *Elementary School Journal* devoted most of a special issue to a report on the Napa study by Stallings and Eileen Krasavage and to responses to it. Yet all the back-and-forth discussion missed the most important question: Why was this small and flawed study the first and only published evaluation of a program that was being used in thousands of schools?

10. *Developer claims that disappointing results are due to poor implementation*. When results fail to support a developer's model, the developer is certain to claim that the model was not properly implemented. This may be true, but it is difficult to prove either way. When a developer can point to well-designed studies in which the method was properly implemented and produced positive effects, the "poor implementation" claim for later studies becomes more supportable. In the case of ITIP, the lack of such studies calls into question the idea that Hunter's methods can be successfully implemented on a broad scale.

11. *Interest in program flags*. At this stage, large numbers of districts move on to other programs, though not usually because of the disappointing evaluations. Nor do programs disappear everywhere or all at once. Districts dropping or no longer emphasizing a program simply begin to outnumber those that are beginning it. This is about where ITIP appears to be today.

12. *Controlled evaluation studies are published*. At long last, controlled evaluation studies, reviews of research, and other articles begin to appear in high-quality research journals. However, the news – usually bad – arrives too late.

STOP THE PENDULUM – I WANT TO GET OFF

In describing the ups and downs of the pendulum of educational change, I've painted a dismal picture. Obviously, the progression of events varies for each innovation, as does the length of time the process takes and the degree to which the innovation takes hold. However, the history of such innovations as programmed instruction and the open classroom follows a pattern similar to that described here. Each program entered widespread use and was already on the wane before controlled evaluations with disappointing findings were published.

The pendulum process outlined above does assume that a program's effects turn out to be disappointing when evaluations are finally conducted. Why is this so often true? One reason is that few educational innovations are designed to insure positive effects in a fair evaluation. To be sure that programs would be effective in fair, controlled comparisons, developers themselves would have to conduct many such evaluations before going public. This rarely happens. However, it is

not clear how much it would matter to the swinging of the pendulum if the program evaluations were positive.

As each innovation swings up and down through the arc of the pendulum, we do learn something that may be of use now or in the future. Nonetheless, if we're going to make generational, beneficial change in education, we're going to have to proceed in a different way.

Two major shifts will have to take place if we are to wean ourselves from faddism and increase the chances for responsible and lasting change in education. First, school districts will have to demand high-quality evaluations of programs before they adopt them. Federal, state, and local governments can assist in this process. Second, school districts will have to focus their staff development efforts not on one-shot workshops, but on extended training and follow-up for a smaller number of programs and practices of proven effectiveness. The emphasis in staff development must shift from scattershot presentations on what's new to systematic implementation of what works.

> The emphasis in staff development must shift from presentations on what's new to implementation of what works.

EVALUATING EVALUATIONS

Educational administrators often complain that they lack the technical sophistication to evaluate the quality of program evaluations. Actually, evaluating evaluations is not so difficult or arcane. In selecting programs for dissemination, school administrators should seek answers to at least the following three questions.

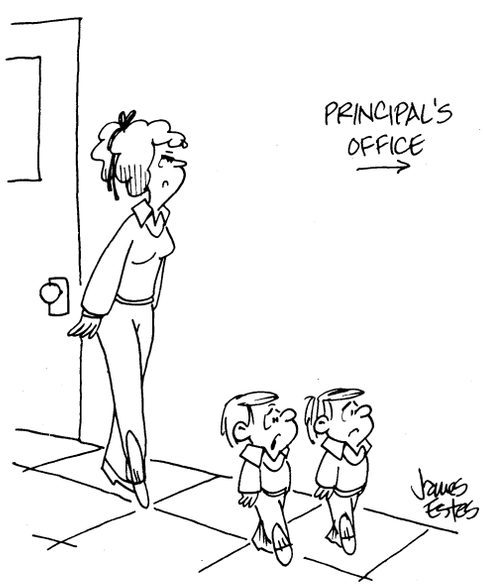
1. *Has the group using the program been compared to a comparable control group?* One hallmark of a high-quality program evaluation is a comparision of schools or classes using the program with similar schools or classes using traditional methods. Ideally, schools or classes are assigned at random to experimental or control conditions; otherwise, it is possible (though much less than ideal) to compare experimental schools to matched control schools. In either case, but especially in the latter, it is crucial to have evidence that experimental and control schools or classes were essentially identical on as many factors as possible at the time of the pretest.

In recent years, rather than using control groups, many educational evaluations have used gains in percentile ranks or in normal-curve equivalents (NCEs) as criteria for program effectiveness. The theory is that a school that tested at the 30th percentile in one year and at the 45th percentile in the following year has obviously improved relative to the norming sample. However, percentile gains or NCE gains have been found to inflate estimates of program effectiveness to an unknown (but often substantial) degree, so that evaluations of this type should be approached with considerable skepticism.[14]

2. *Did the posttest assess objectives that were being pursued equally by experimental and control classes?* Sometimes program evaluations look very positive because the program pursues nontraditional objectives and then assesses achievement on those objectives. For example, evaluations of IBM's Writing to Read program typically assess the reading and writing skills of kindergartners and first-graders. The writing assessments are obviously biased in favor of the program because writing is not traditionally taught at these grade levels. The reading measures at the kindergarten level are often biased because many traditional kindergarten programs that are used as control groups for Writing to Read are nonacademic in emphasis. Similarly, studies of "curriculum alignment" programs or of other programs (such as mastery learning) that include an element of curriculum alignment inherently give the experimental program an advantage, because by their nature such programs focus students on a more limited set of objectives not pursued to the same degree in traditional programs.[15]

3. *Was the program evaluated under realistic conditions over realistic time periods?* Programs are often evaluated under highly artificial conditions or for brief time periods. Before adopting a program, administrators should be sure that the program has been evaluated in real schools somewhat like their own for at least a full school year. If the developer's positive evaluations have been replicated by others, administrators can be even more confident of the effectiveness of the program.

As a final safeguard against adopting ineffective programs, school districts that intend to adopt new programs on a broad scale should first conduct evaluations of those programs on a smaller scale, com-



"I don't understand it. They usually crumble when I turn on my boyish charm."

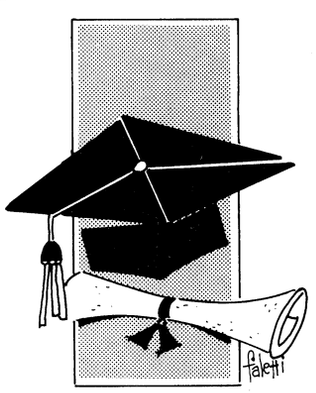
> If faddism in education is ever to end, decisions about adopting or maintaining programs must be based on reliable data.

paring experimental and control schools or classes on fair measures over extended time periods.

GOVERNMENT SUPPORT FOR EVALUATIONS

If faddism in education is ever to end, decisions about adopting or maintaining programs must be based on reliable, widely respected data. Federal or state governments can help in several ways.

First, the federal government could provide a clearinghouse for program evaluations. Such a clearinghouse could serve as a central source of data, particularly for evaluations conducted by school districts. It is probable that, among the 16,000 U.S. school districts, hundreds of top-quality evaluations are known only to a few people within the district. Unsuccessful evaluations are particularly unlikely to be disseminated beyond district boundaries. At minimal cost and effort, the U.S. Department of Education could encourage districts to simply mail copies of such evaluations to a central clearinghouse. For evaluations that are federally funded, certain standards could be required as well.



The Department of Education's Joint Dissemination Review Panel (JDRP), which reviews evaluations of programs developed with federal funds, already plays a similar role. Programs that pass JDRP review are eligible for dissemination through the National Diffusion Network. However, the JDRP process is flawed in two major ways. First, the developers are the only sources of the evaluations, and they obviously have a stake in presenting the data in a positive light. Second, most JDRP-approved programs fail to use experimental/control comparisons; most use NCE or percentile gains which, as noted earlier, tend to overstate program effects.

Second, either as part of their funding of educational innovation or in a separate grant program, federal or state governments could contract with school districts to conduct top-quality evaluations of promising practices. The dissemination of the findings should be included in the contract.

Third, the best of all approaches would be a system of independent evaluation laboratories, funded by the state or federal government to conduct evaluations of promising programs and disseminate the findings. This could provide widely respected and believable results that districts could hardly ignore in making decisions about program adoption.

If we had a set of programs that were proven effective, with effects that were beyond dispute, we could turn our attention to effective implementation. Serious educational change takes time and money; to move beyond faddism we need to invest in a small number of proven programs and make sure that they are properly implemented and are making the differences they should make. Instead of running one-shot workshops, school districts would provide extensive training, classroom follow-up, peer coaching, or even in-school program facilitators to help make the transition from traditional methods to more effective programs. If we know that a program is effective when properly implemented, it is worthwhile to stick with the program for as long as it takes to make it effective. In this respect, Hunter has made an important contribution: in implementing her model, many school districts have for the first time engaged in long-term, serious training and follow-up.

Faddism is so well entrenched in American education that uprooting it will take time and concerted effort, probably with significant government involvement. However, if education is ever to make serious generational progress, educators must somehow stop the pendulum by focusing their efforts to improve education on programs that are effective, rather than on those that are merely new and sound good. Otherwise, we will endlessly repeat the process that led us in and out of the open classroom, in and out of individualized instruction, and in and out of Madeline Hunter's ITIP.

---

1. Madeline Hunter, "Comments on the Napa County, California, Follow-Through Project," *Elementary School Journal*, vol. 87, 1986, pp. 173-79.
2. Madeline Hunter, *Teach More – Faster!* (1969; revised ed., El Segundo, Calif.: TIP Publications, 1980).
3. Jane Stallings, "A Study of Implementation of Madeline Hunter's Model and Its Effects on Students," *Journal of Educational Research*, vol. 78, 1985, pp. 325-37.
4. Jane Stallings and Eileen M. Krasavage, "Program Implementation and Student Achievement in a Four-Year Madeline Hunter Follow-Through Project," *Elementary School Journal*, vol. 87, 1986, pp. 117-38.
5. Robert E. Slavin, "The Napa Evaluation of Madeline Hunter's ITIP: Lessons Learned," *Elementary School Journal*, vol. 87, 1986, pp. 165-71.
6. Pamela Robbins, personal communication, 6 May 1987.
7. Hunter, "Comments on the Napa County. . . ."
8. John D. Tudor, "Background for the South Carolina Implementation of the PET Inservice Teacher Training Program," paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1988.
9. Joan Assey and Elaine Grieb, "The Program for Effective Teaching as Implemented in Richland County School District 2," paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1988.
10. Janelle Rivers, "Teacher Perceptions of PET with Special Emphasis on Coaching," paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1988.
11. Garrett Mandeville, "An Evaluation of PET Using Extant Achievement Test Data," paper presented at the annual meeting of the American Educational Research Association, New Orleans, April 1988.
12. Garrett Mandeville, personal communication, 14 June 1988.
13. James F. Donovan, David A. Sousa, and Herbert J. Walberg, "The Impact of Staff Development on Implementation and Student Achievement," *Journal of Educational Research*, vol. 80, 1987, pp. 348-51.
14. Robert E. Slavin, "Making Chapter 1 Make a Difference," *Phi Delta Kappan*, October 1987, pp. 110-19; and Robert E. Slavin, Nancy L. Karweit, and Nancy A. Madden, eds., *Effective Programs for Students at Risk* (Boston, Mass.: Allyn and Bacon, 1989).
15. Robert E. Slavin, "Mastery Learning Reconsidered," *Review of Educational Research*, vol. 57, 1987, pp. 175-213.