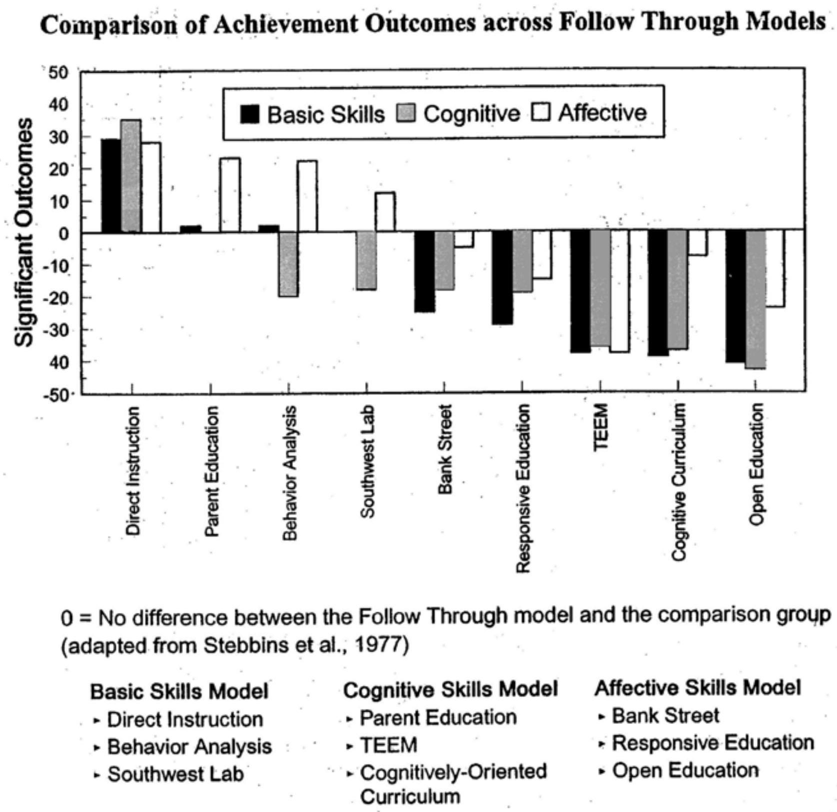
Summary of Presentation to Council of Scientific Society Presidents, December 8, 2002

This summary presents data on the effectiveness of Direct Instruction, indicates what I believe to be my most seminal achievement, and concludes with observations about the power in supplementing basic research with detailed evaluations and study of programs and interventions that work most effectively.

Data:

Figure 1 shows the overall advantage of Direct Instruction over the other eight major models tested in Project Follow Through, which was the largest educational experiment ever conducted. The project was implemented in more than 100 communities, starting in 1969. The models ranged from direct approaches that focused on what the children were to learn to cognitive orientations (based on developmental theories and practices) and models that set out to control the “affective,” self-image variables.

Each district selected one model and received $750 per student, in addition to the district’s maintenance of effort, to fully implement the model of its choice. In addition each sponsor was provided with funds sufficient to train and monitor each site. Follow Through was evaluated by Stanford Research Institute and Abt Associates in 1977. The evaluation involved the last wave of 10,000 at-risk third-graders who had gone through the nine models.

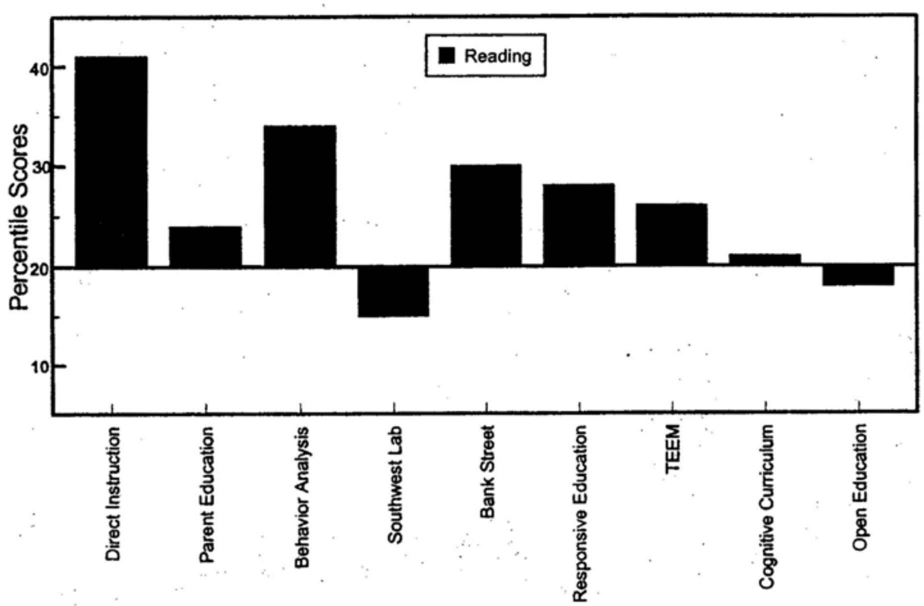


1

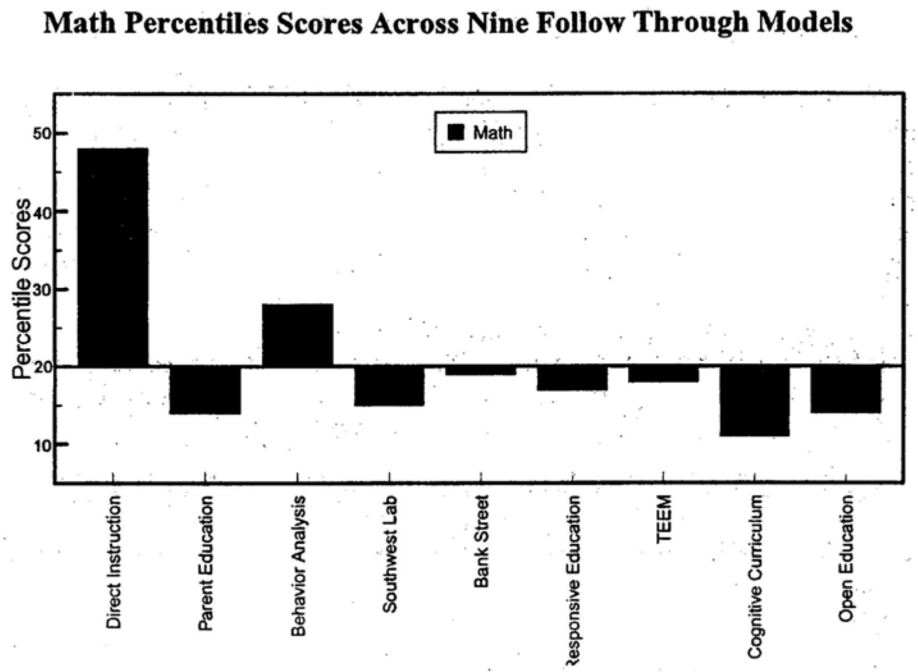
Figure 1 shows the net educationally significant differences between each model and the comparison children who went through extant Title One programs (both a pooled comparison group and one based on the characteristics of each school). Differences were considered educationally significant if they were at least one-fourth standard deviation. Direct Instruction (DI) had the highest number of net educationally significant differences in basic skills, cognitive skills, and affective measures (how much the children took responsibility for their successes and failures). As the figure shows, most of the models, which parallel models that are used extensively in schools today, performed far worse than the comparison children.

Figures 2, 3, and 4 show the percentile scores for the nine models in Reading, Math, and Language. The baseline is the 20th percentile, assumed to be the standard for Title-One students. DI outperformed the other models by a substantial margin in all three subjects (and in spelling). The differences are more striking in math and language, with most of the models performing at or below the 20th percentile.

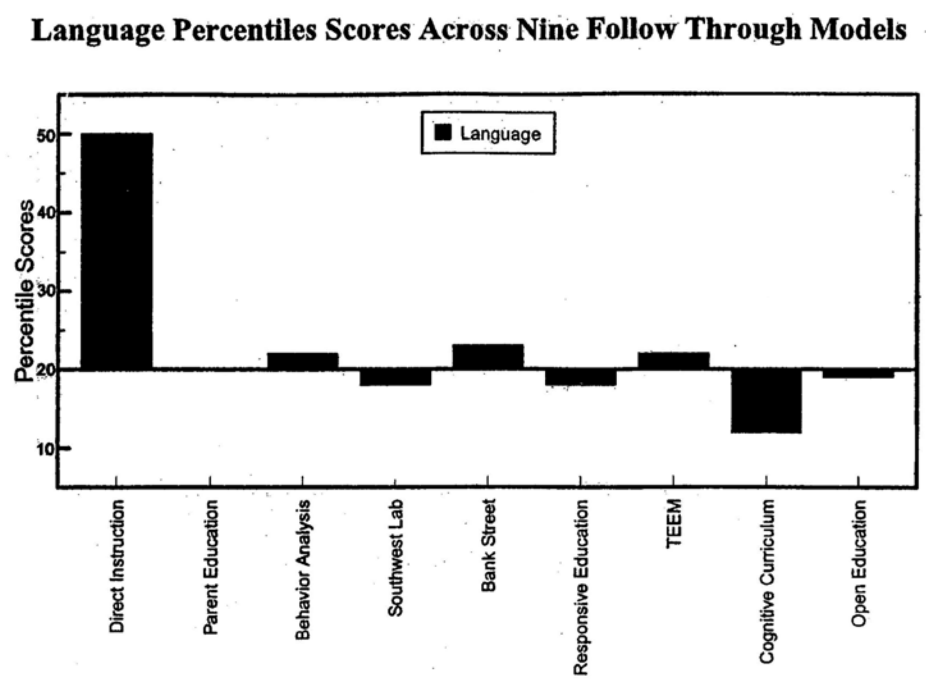
Reading Percentiles Scores Across Nine Follow Through Models



2

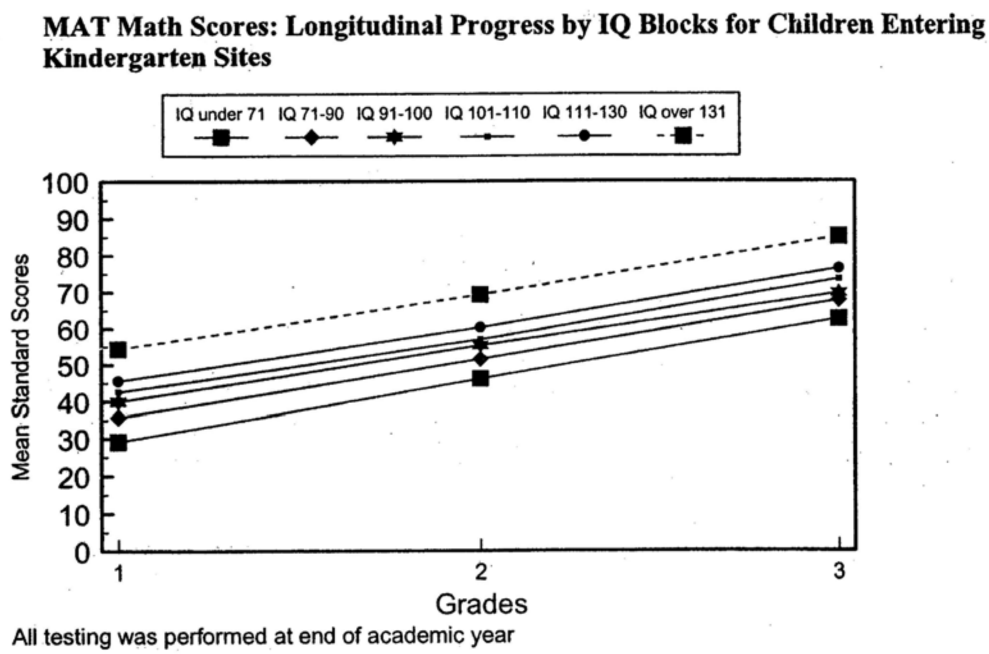

Math Percentiles Scores Across Nine Follow Through Models
Math
Percentile Scores
50
40
30
20
10
Direct Instruction
Parent Education
Behavior Analysis
Southwest Lab
Bank Street
Responsive Education
TEEM
Cognitive Curriculum
Open Education

3


Language Percentiles Scores Across Nine Follow Through Models
Language
Percentile Scores
50
40
30
20
10
Direct Instruction
Parent Education
Behavior Analysis
Southwest Lab
Bank Street
Responsive Education
TEEM
Cognitive Curriculum
Open Education

4

Figure 5 shows the math performance (in standard scores) of DI students of various IQ ranges (from over 131 to under 71) as they progressed from grade 1 through grade 3. The lines are parallel, suggesting that the same rate of the students in achievement was realized for all students. This outcome is partly an artifact of the priorities of the DI model, because it focused disproportionately on the lower performers, those students less likely to succeed. With more emphasis on the higher performers, their performance could have been accelerated more dramatically.



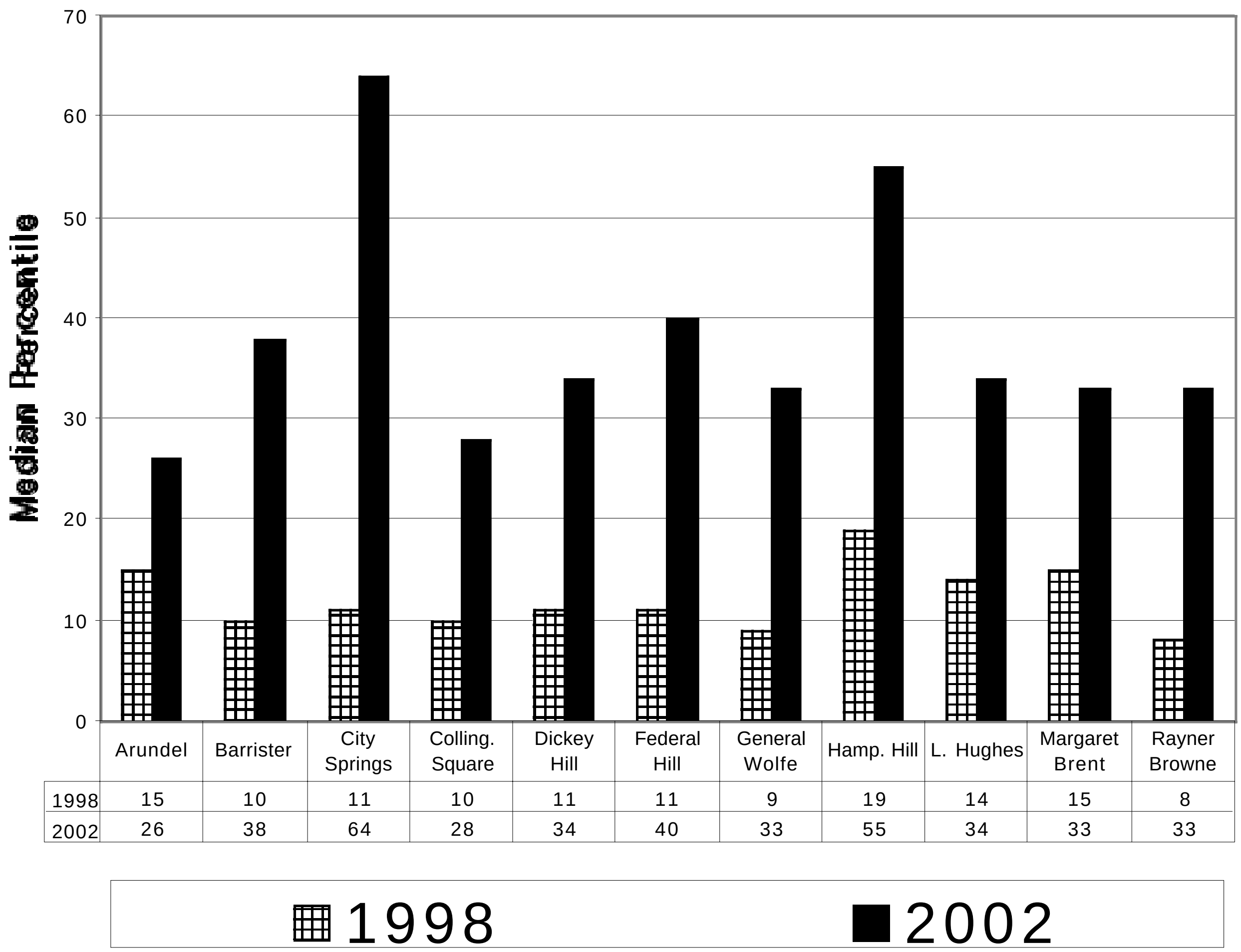
5

More-current data on DI is presented through figures 6-11. Figures 6-9 show the performance of students in a DI implementation managed by The National Institute for Direct Instruction. Figures 10 and 11 show the results of an independent organization, the RITE Institute in Houston, Texas.

The Baltimore data compares performance of the 11 schools NIFDI worked with since 1998. Figure 6 shows fifth-grade reading in percentiles. Although there is great variation in the 2002 performance of the sites, nearly all of them more than doubled their percentile scores, with one of them going from the 11th to the 64th percentile. The variation is largely a function of the extent to which the school followed the NIFDI procedures for implementing, rather than the district guidelines, which are not consistent with those of NIFDI. The highest performing school followed the NIFDI guidelines with the greatest fidelity.

Figures 7 and 8 show first-grade and fifth-grade math performance. Most of the first-grade percentiles in 2002 are over the 60th percentile and all are more than double the 1998 level. The fifth- grade performance is not nearly as high, and not as high as the 5th-grade reading

CTBS Reading Scores in NIFDI Baltimore Schools GRADE 5

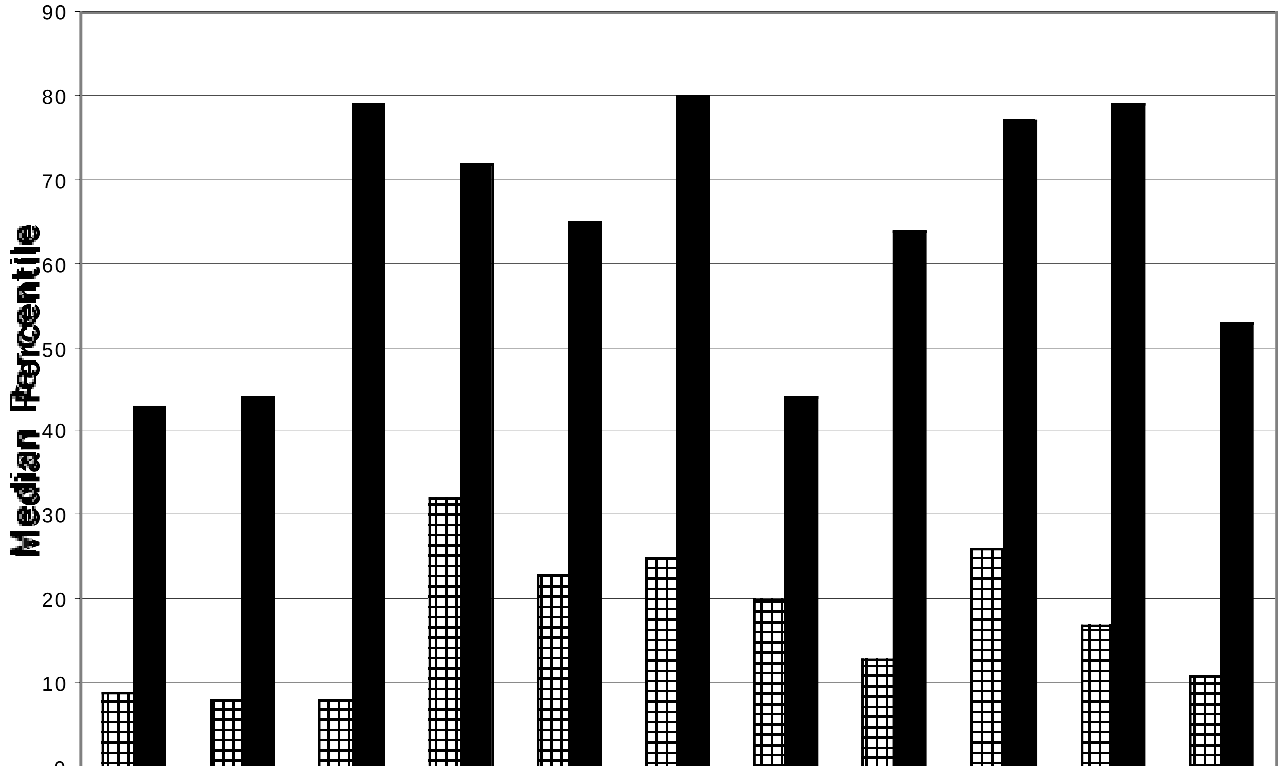


	Arundel	Barrister	City Springs	Colling. Square	Dickey Hill	Federal Hill	General Wolfe	Hamp. Hill	L. Hughes	Margaret Brent	Rayner Browne
1998	15	10	11	10	11	11	9	19	14	15	8
2002	26	38	64	28	34	40	33	55	34	33	33

performance. This difference is largely a function of turnover. Although a high turnover rate (25% percent per year) affects reading, its results create greater stress in math in grades 4 and 5. The reasons are: (1) it is harder to train teachers in these grades to teach math effectively; and (2) the likelihood of incoming students fitting into any group of students that has been in the school since K or 1 is zero. (Reading is different because at least some students coming in read well enough to perform in one of groups in the fourth or fifth grade.) To accommodate incoming students for math, the higher performers (continuing students) are often compromised because they represent a relatively small group, and there is no practical way to place them appropriately.

Even so, the 5th-grade math performance is far superior to what it had been in 1998, with nine of the schools at least doubling their 1998 scores.

CTBS Math Scores
NIFDI Baltimore Schools
GRADE 1

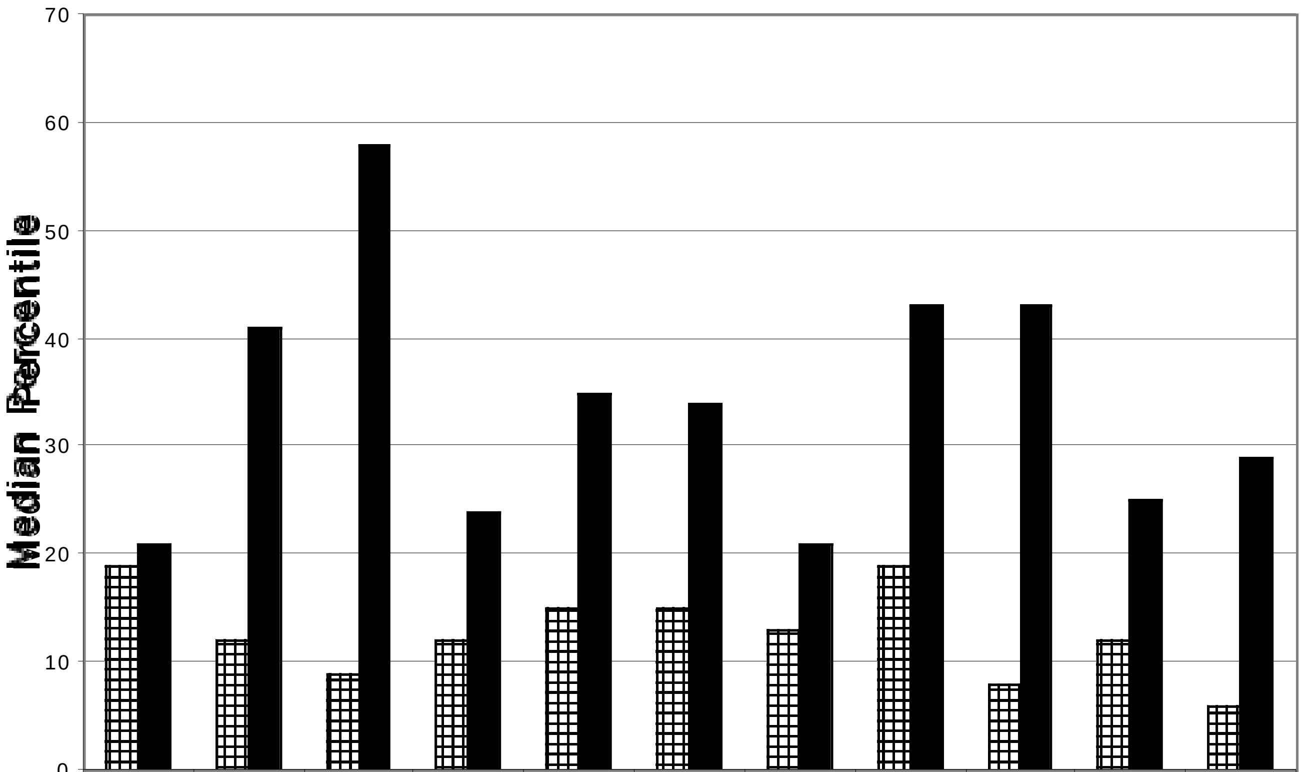


	Arundel	Barrister	City Springs	Collington Square	Dickey Hill	Federal Hill	General Wolfe	Hamp. Hill	Langston Hughes	Margaret Brent	Rayner Browne
1998	9	8	8	32	23	25	20	13	26	17	11
2002	43	44	79	72	65	80	44	64	77	79	53

1998 2002

7

CTBS Math Scores
NIFDI Baltimore Schools
GRADE 5



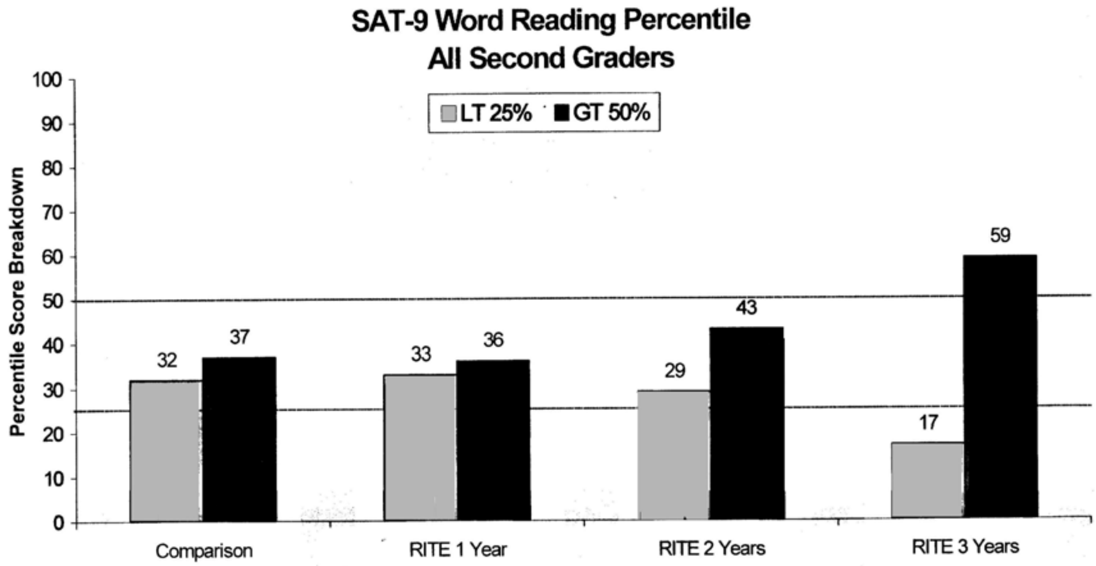
	Arundel	Barrister	City Springs	Collington Square	Dickey Hill	Federal Hill	General Wolfe	Hamp. Hill	Langston Hughes	Margaret Brent	Rayner Browne
1998	19	12	9	12	15	15	13	19	8	12	6
2002	21	41	58	24	35	34	21	43	43	25	29

1998   2002

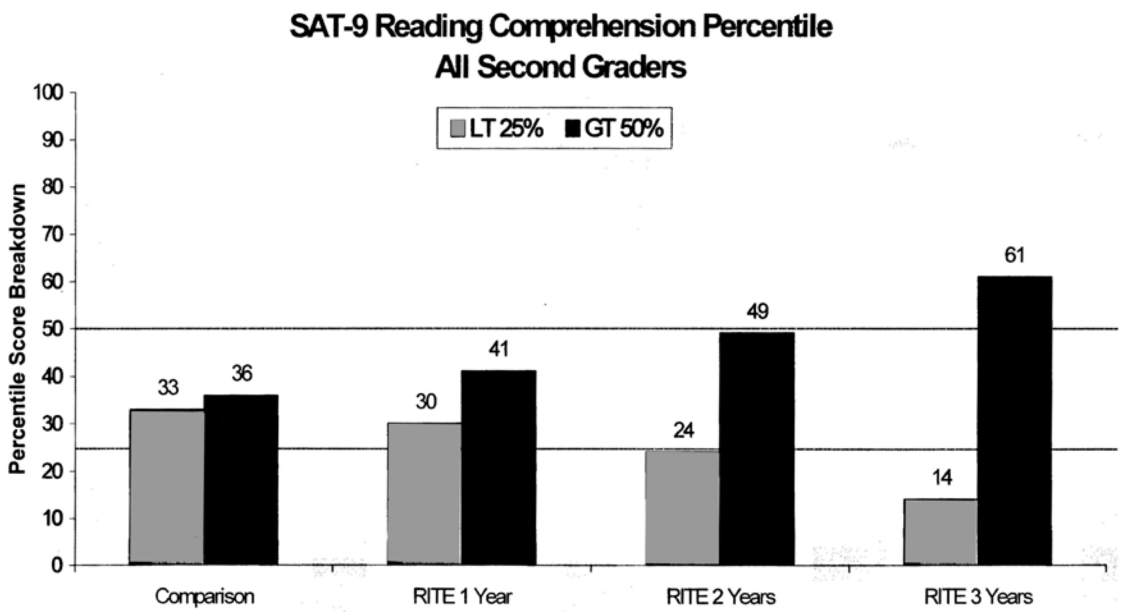
8

Figures 9 and 10 show that there is significant stress on an effective program that teaches reading to at-risk children. The figures are based on 4,000 children who went through the RITE program, which uses the same implementation guidelines as those for Follow Through and Baltimore, and about the same number of comparison children. The comparison schools used what is considered to be a superior, research-based phonics program. The figures show that the performance of second-graders in the RITE program (on the Stanford Achievement Test) increased as a function of the number of years they were in the program, both in word reading and reading comprehension. The 1-year RITE students (second grade only) perform about the same as the comparison children. Those in the RITE program for two years perform close to the test “norm,” with close to 50% of the students above the 50th percentile and close to 25% below the 25th percentile. The 3-year RITE students are above average, with only 17% of these students performing below the 25th percentile.

Because the data on virtually every DI site in Follow Through, Baltimore and Houston have more than 25% annual turnover, the potential for accelerating the performances of at-risk students is quite a bit greater than the composite results indicate.


SAT-9 Word Reading Percentile
All Second Graders
LT 25%
GT 50%
Percentile Score Breakdown
100
90
80
70
60
50
40
30
20
10
0
32
37
33
36
29
43
17
59
Comparison
RITE 1 Year
RITE 2 Years
RITE 3 Years

9


SAT-9 Reading Comprehension Percentile
All Second Graders
LT 25%
GT 50%
Percentile Score Breakdown
100
90
80
70
60
50
40
30
20
10
0
33
36
30
41
24
49
14
61
Comparison
RITE 1 Year
RITE 2 Years
RITE 3 Years

10

My most seminal achievement:

I believe it is the systematic way we approach the problem of designing effective interventions, and, more specifically, effective instructional sequences. We try to proceed scientifically, in a way that generates data, and we try to live by the data.

Figure 11 shows the major implementation variables that we attempt to control. Two important points about these variables are: (1) Some of the variables are not evident unless a program is well implemented; (2) the variables interact with each other and with some that are not listed but implied (such as teacher training).

The Controllable Variables:	**The Instructional Program:**
Effective instructional programs	Introduces only 10% new each lesson
Homogeneous instructional groups	Provides teaching on new skills for 3 consecutive lessons
Schoolwide participation	Teaches all the component skills needed for later applications
Placement based on 70% correct	Has 6-10 ongoing tracks (topics within each lesson)
Adequate schedules	Provides for the continued use of everything taught
Coordination of schedules	Provides for a high rate of student-teacher evaluation
Teaching to mastery	
Schoolwide reinforcement-management procedures	
Problem-identification and problem-solving provisions.	
Ongoing, current, student-teacher data.	

11

The assumption that generates the different variables is that teaching at-risk students is a desperate enterprise, with respect to the rate at which things must be taught. Unless children who historically have learned at a slower rate than their more advantaged peers receive instruction that accelerates their performance, they will never catch up. So the program that succeeds must be scrupulously efficient and intolerant of either teacher or student failure.

Effective instructional programs. These are the *sine qua non* of an implementation. The program is the vehicle that communicates specific content to the children, guides the teachers, and sets the stage for projections about what is possible with a given group of children during the upcoming school year. If the program is not well designed, sufficiently standardized, and “trainable,” many of the other variables become moot.

Homogeneous instructional groups are necessary so that each individual in the group will benefit maximally from the teacher’s presentation. If the group is composed of children who already know the material and those who are so lacking in prerequisite knowledge that they couldn’t possibly learn the material in the allotted time, the program is not effectively accelerating the performance of all students.

Schoolwide participation is necessary so that it is possible to make projections about what each instructional group will have learned in each subject by the end of the year, and where the individuals would be best placed at the beginning of the next year. We must know what all teachers are teaching so that we achieve maximum efficiency in “interchangeable parts,” (addressing problems of teacher replacement and student transfers).

Placement based on 70% correct. If the instructional programs are designed so that they are progressive and teach so much new material each lesson, while reviewing what had been presented earlier, students who perform at 70% correct on any skill or content introduced for the first time are properly placed. The percentage implies that the child is in a position to master the new material today. (If the child is only about 50% correct on the tasks the teacher presents while teaching the new material, the child must learn too much in the allotted time.)

Adequate schedules and coordinated schedules. Unless there are provisions for adequate amounts of daily instruction in all subjects, performance will suffer. We can't spend too much time on reading. We can't have elastic schedules. Schedules must be followed religiously. Furthermore, they must be coordinated so that it is possible to deploy aides to various classrooms at times when specific instruction is occurring.

Teaching to mastery. After each lesson, all children should be virtually 100% correct on all material presented during the lesson. This objective is not readily attainable unless the group is homogeneous in performance, the schedule followed, the children placed at 70% correct on new material.

Schoolwide reinforcement-management procedures. These involve rules of behavior for the various school activities. More broadly, the system must be designed to celebrate academic achievements, and to provide students with information about their role in this school and with data that they are capable of achieving in the academic arena. Performance is recognized, not just through an end-of-year awards ceremony, but weekly through student-performance progress summaries.

Problem-identification and problem-solving provisions. Those provisions rely on ongoing, current student-teacher data. The only way the school will improve is to do things better than it does now. Therefore, the emphasis must be on problems, things that need fixing. The purpose of identifying problems is not to "evaluate" or "shame" but to make it possible to solve the problem in a timely manner. Therefore, basic monitoring and data-collection procedures are needed to show the progress of each child. Furthermore, the program is designed so that a coach or trainer would be able to walk into any classroom, refer to the posted data, and observe the responses of the children as they are being instructed and identify any problems in placement of individuals, rate of progress for the group, teaching problems involving mastery, and problems with time management. The precise identification of these problems in a timely way provides straightforward implications about how to solve them.

The instructional program. The features of the program interface with the broader implementation variables. The program has a quite-different design than that of the traditional program. It does not devote an entire period to a particular topic. Rather, the program is designed like a flight of stairs that introduces only about 10% brand new material in the current lesson. It reviews the 10% introduced in the last lesson and the 10% in the lesson before that one. The program does not assume that anything is taught in less than three consecutive lessons. Therefore, about 30% of the lesson is not assumed to be thoroughly learned. The remaining 70% of the lesson consists of earlier-taught material that is either being reviewed or that is incorporated in applications. This provision dovetails with the 70% first-time correct requirement for placing children in the program.

Each lesson consists of 6-10 ongoing tracks (topics), each independent of the others, but later combined to create new operations. The program is designed so it teaches all the component skills that are needed for later applications. It also provides for the continued use of everything that is taught. Finally, it is designed so that there is a high rate of student responses, sometimes in the form of group, unison responses. The high rate of responding is not merely a means by which the skills are made more automatic and thoroughly learned; they also provide a high rate of information about student performance for the teacher. Specific problems that children have with the material are made available to the teacher in a timely way.

The design of the programs involves the application of analytical principles, empirical rules of thumb, and field-test data on student and teacher performance. The analytical principles generate procedures for designing sets of examples that show how things are different, how they are the same and, most of all, how to construct a series of examples that are consistent with only a single interpretation.

For example, in teaching +1 facts in math, the program would not always present the facts in the counting order, 1+1, 2+1, 3+1, because the series generates two interpretations or strategies that lead to correct answers. (1) You must listen to each fact. (2) You must attend to the last answer and say the next counting number, without attending to the details of any problem. Because both strategies will lead to correct answers, the procedure must be modified or supplemented so that interpretation 2 is ruled out and the child attends to the details of each fact.

The empirical rules of thumb serve as guidelines for mapping out the number of lessons required to teach a particular skill (given the assumption that on no lesson will the presentation of the new skill involve more than 30% new material that may not be mastered).

The final step in the process is field-testing in a way that will reveal possible flaws with the program. This testing does not involve a large number of classrooms, but typically at least a couple of naive teachers (those who have never taught DI before) and at least one reliable DI teacher who understands that the game is to follow the program as written (even though the teacher may be quite aware of how to remedy a problem she observes). Ideally, there will be more than one field-testing of the material; a second group (or teacher) presenting the sequence about a month behind the others. This group permits adjustments to be made as soon as they are revealed with the original tryout groups, and for the adjustments to be tested.

The teacher records the starting and stopping time for each exercise in the lesson and indicates the amount of time the students spent on the independent work. (The programs provide for very little homework. The assumption is that all the skills the children need are best taught in class, where there is greater control.) The teacher also indicates any problems that were encountered in each lesson. These do not have to be detailed, simply the number of any tasks that presented the problem and a note like, “Hard,” “many mistakes”, “awkward,” or “wordy.” All student written work is analyzed for mistakes. If more than about 18% of the students miss any item, we analyze the item, try to identify the problem, and provide a remedy.

The assumption throughout the initial development and modification of the instructional sequence is that if the children or teachers have problems presenting the material, the program is the cause of the problem and the program is to be changed. This assumption does not flatter one's ego because it has happened that we have worked on a level of a program for over a year and, following an initial field tryout that disclosed far too many problems with the sequence, had to throw out virtually the entire program. The only solace is that we try not to make the same mistakes again. In no small way, whatever skill and knowledge we have acquired is a function of the feedback teachers and students have provided as they go through programs or procedures we develop. Their behavior shows us whether the sequence is adequate. By revealing specific problems, their behavior also implies the nature of the changes that must occur if the program is to be effective.

Basic research and identification of highly effective programs:

Because not all the variables of highly effective teaching communications are obvious by inspecting either mediocre programs or excellent programs that are poorly implemented, the study of highly effective programs that are fully implemented will add greatly to the knowledge base of educational research. The formula for this research would be to implement a program or model so that it met all the developer's guidelines. Systematic manipulation of single variables would then disclose the relative importance of the variable (the percentage of the variance it accounts for), how it interacts with other variables, and whether the variable as originally conceived by the developer is relatively efficient (with respect to amount of exposure) and effective (with respect to student-teacher performance).

The recent formation of the Institute of Education Sciences with its emphasis on applied research—what works and relatively how well it works—should have a salutary effect on the field and should provide the consumers of instructional material (the district and state) with the kind of endorsements they need to acquire new attitudes both about what works well and why programs with the potential to work well must be implemented with fidelity. (We don't buy high-performance cars and then run them with two flat tires.) The result will not only be a great acceleration in the development of highly effective programs but far more effective interventions for those children who, without such programs, will not be competitive and will not be able to make choices about whether they will go to college. The question of their potential choices when they are 17 will usually have been answered by the time they reach the fourth grade.

References

A discussion of Follow Through and other DI data through 1995 appear in:
Adams, G. L., & Engelmann, S. (1996). *Research on Direct Instruction: 25 years beyond DISTAR.* Seattle, WA: Educational Achievement Systems.

The data on Baltimore have not been published, but are available online at: www.bcps.k12.md.us/terranova/terraNova.html.

Data on the RITE Program appear in:
Carlson, C. D., Francis, D. J., Tatif, L., Priebe, S., & Ferguson, C. (2001-2002). *RITE program external evaluation 2001-2002.* Houston, TX: The Texas Institute for Measurement, Evaluation, and Statistics. On-line: www.hlsr.com/rite.

The analytical principles involved in basic program design are articulated in:
Engelmann, S., & Carnine, D. (1991). *Theory of Instruction: Principles and applications.* Eugene, OR: ADI Press.