ADI NEWS

VOLUME 8, NUMBER 4 THE ASSOCIATION FOR DIRECT INSTRUCTION SUMMER, 1989

SPECIAL HISTORICAL ISSUE NO. III

Join the Association for Direct Instruction

ADI

Membership Options:

A) Regular Membership: $15.00 per year (includes one year of *DI NEWS* and a 20% discount on ADI sponsored items and events).

B) Student Membership: $7.00 per year (includes one year of *DI NEWS* and a 40% discount on ADI sponsored items and events).

C) Sustaining Membership: $30.00 or more per year (includes regular membership privileges and recognition of your support in the *DI NEWS*).

D) Institutional Membership: $50.00 per year (includes 5 subscriptions to the *DI NEWS* and membership privileges for 5 staff people).

E) *DI NEWS* Subscription only: $7.00 per year (outside North America and Hawaii $10.00 per year).

ADI sponsored products and events include books and other materials published or marketed by the Association for Direct Instruction. The ADI NEWS is published 4 times a year (Fall, Winter, Spring and Summer).

To join the Association, complete the bottom portion of this form and mail it, with your check in U.S. funds to ADI.

Check one:
I wish to become an Association member. Please enroll me as a:

_______ Regular Member ($15.00 annually)

_______ Student Member ($7.00 annually)

_______ Sustaining Member ($30.00 or more annually)

_______ Institutional Membership ($50.00 annually)

_______ I wish to subscribe to the DI NEWS only ($7.00 annually; $10.00 outside North America & Hawaii)

Name: ______________________________

Address: ______________________________

City, State, Zip: ______________________________

Phone: ______________________________

School District or Agency: ______________________________

Position: ______________________________

The **Direct Instruction News** is published Fall, Winter, Spring and Summer, and is distributed by mail to members of the Association for Direct Instruction. Membership and subscription information may be found on page 41 of this newsletter. Readers are invited to submit articles for publication relating to DI. Send contributions to: The Association for Direct Instruction, P.O. Box 10252, Eugene, Oregon, 97440.

From the Editor:

This is the third and final (for now) historical issue of the *ADI News.* With these historical issues we have brought together most of the important articles from the first 7 years of the *ADI News* and published them in the new magazine format that should be easier to file and save for future reference.

In this issue we look at issues in implementation and management of educational change, additional research studies related to DI that may have enduring interest, and three review articles touching on controversies in reading and the effectiveness of DI.

The future of *ADI News* will in part depend on your contributions. Please help us find good articles through your personal research and thoughts, and through sending us suggestions for articles we should reprint.

Wes Becker
Editor

Large City Implementation Study

by Russell Gersten
Leslie Zoref
Doug Carnine

From *ADI News* Volume 1, Number 1 (Fall, 1981)

In August 1978, a large urban school district contracted with the University of Oregon DI Follow Through Project to supervise the implementation of the DI Model in 10 kindergarten and 11 first grade Follow Through classrooms. The DI Model represented a very different educational philosophy and system than the one that had dominated the previous 10 years of Follow Through in that city.

One month later, the U.S. Office of Education funded a 2-year implementation study of the DI Model to better understand change processes in a large urban school district. The study had two goals:

1. To document, via interviews and case study, the process of adoption of a highly structured educational model in a community which had previously used a "loose coupling" laissez-faire approach.
2. To develop valid and reliable measures of implementation of the DI Model and to assess to what extent measures of implementation correlate with achievement gains in Reading and Math.

Primary responsibility for the first objective was subcontracted to an independent agency.

Impact of DI on the Community

Despite an extraordinary amount of turmoil, the first year of implementation was considered a success in terms of student achievement. Reading scores for kindergarten and first grade were higher than they had ever been. (Math was not implemented until the second year.) Early in the second year of the program, a judge, who had been overseeing the city's voluntary desegregation program, ordered an examination of the quality of education in the 28 minority schools. This was part of a slow, but growing trend in busing decisions, in which both the quality of teaching and the racial composition of schools are considered. He appointed three educators from outside the district to observe educational practices, and then recommended procedures likely to improve the quality of education. The seven Follow Through schools were among the schools targeted. The court-appointed team observed that typical educational practices in the minority schools were weak. They cited the lack of active instructional leadership by the principals in these schools, and the tendency to blame poor academic achievement of students on factors such as poverty, ethnicity, and lack of parental interest, rather than seriously looking at the instructional system. They also criticized the extensive use of "pull out" programs for remedial reading and the consequent diffusion of responsibility for student growth. Finally, the absence of intensive, high-quality, concrete inservice training in minority schools was cited. They found two exceptions to this generally glum picture—the seven schools with Direct Instruction Follow Through model and two schools with a bilingual program. They praised the high-quality of concrete, down-to-earth, technical assistance offered by the supervisors, and the high-level of time students spent actively engaged in reading, math and language activities.

The judge's decision reflected the findings of the team, and his own investigation of the East St. Louis DI Follow Through program. Here are some excerpts:

> *Recent experience with schools in poor minority urban districts reveals that if the schools are properly run, virtually all students, no matter how disadvantaged, can be taught to read, write and calculate at a level sufficient to function in American Society...*
>
> *East St. Louis, Illinois, began a Follow Through program in 1968 using Distar. By remaining with the program they have produced outstanding results. The pupils of this district, which is over 90% black, test on pre-school examinations below 88% of the nation's students, but by the end of the third grade, they test above 50% of the nation's students in both reading and mathematics...*
>
> *In the past school year, the mean reading score in Distar kindergartens was at the 72nd percentile (range form 60th to 87th percentile). For mathematics, the mean total score was at the 57th percentile. In first grade, the children were tested at the 18th percentile in reading and 20th percentile in mathematics at the beginning of school, and they tested at the 46th percentile (average) at the end of the year in both reading and mathematics. There were seven classes combined in these averages.*

He goes on to discuss the proven superiority of phonics-based reading program, the necessity for a system to monitor implementation of the educational program, the need for concrete supervisor feedback, and concluded with the following:

Traumatic though it may be to the community, busing is an easy way out—temporarily. Busing can be carried out by superintendents reluctant or unwilling to do so. It is visible, easily enforced by the Court and immediately satisfying to those minorities who see it as a symbol of victory over the white community or who have been misled to believe that is will magically produce better educational results for their children.

Educating children requires willing cooperation on the part of the top administrators, principals, and teachers and the organizations that represent them. It requires hard work, inspiration, imagination and perseverance. It is less newsworthy and less visible, but more fulfilling. It works to lift the children out of isolation—permanently!

The DISTAR Follow Through programs have received extensive coverage with a total of nine articles appearing in the city's newspapers during the summer of 1980, when the district was deciding on expansion into other schools. A lead editorial was called "DISTAR's Success Story". Here are some excerpts:

Improving the quality of education available at 23 predominantly black and hispanic schools would be a moral imperative even it it were not also essential to the success of this city's school integration effort.

The parents of white, middle-class youngsters will never be persuaded to send their children to minority schools so long as those already enrolled in these schools continue to score abysmally low in tests of basic reading and mathematical skills.

More importantly, failure to enhance educational opportunities for minority children only condemns them to the tragic patterns of the past, including sharply diminished employment and career prospects.

Fortunately, the problem is not insoluble. Help is available in the form of a classroom program known as Direct Instruction, or DISTAR.

DISTAR's stress on rote learning and regimentation [note the common misinterpretation] offends many school administrators and teachers who favor a more "creative," less structured classroom atmosphere. But that approach, typical of middle-class schools, has proved a disastrous failure for minority children from poverty backgrounds. Indeed, it has contributed to the decline in test scores among middle-class children.

Conversely, Direct Instruction has achieved results far surpassing any of the other 15 or so experimental techniques tested in minority schools around the country since the late 1960's. DISTAR delivers what other methods only promise; namely, raising the reading and math scores of disadvantaged children to national norms for their grade level. That being the case, it seems absurd to oppose DISTAR on grounds of arbitrary and discredited philosophical bias.

The legacy of that bias is a conviction shared by all too many educators that minority children simply cannot learn at rates anywhere near those commonly achieved by the middle-class majority. That pernicious myth is challenged eloquently by the Rev._____ in an interview published opposite this page.

The city school board is scheduled to vote Tuesday on whether to expand its DISTAR experiment to 18 additional classes. Superior Court Judge______, who is overseeing the district's court-ordered voluntary integration plan, has indicated he favors a wider implementation of the Direct Instruction method.

So do we. And we trust that a majority of the school board will agree when it meets on Tuesday. [Reproduced with permission of the San Diego Union.]

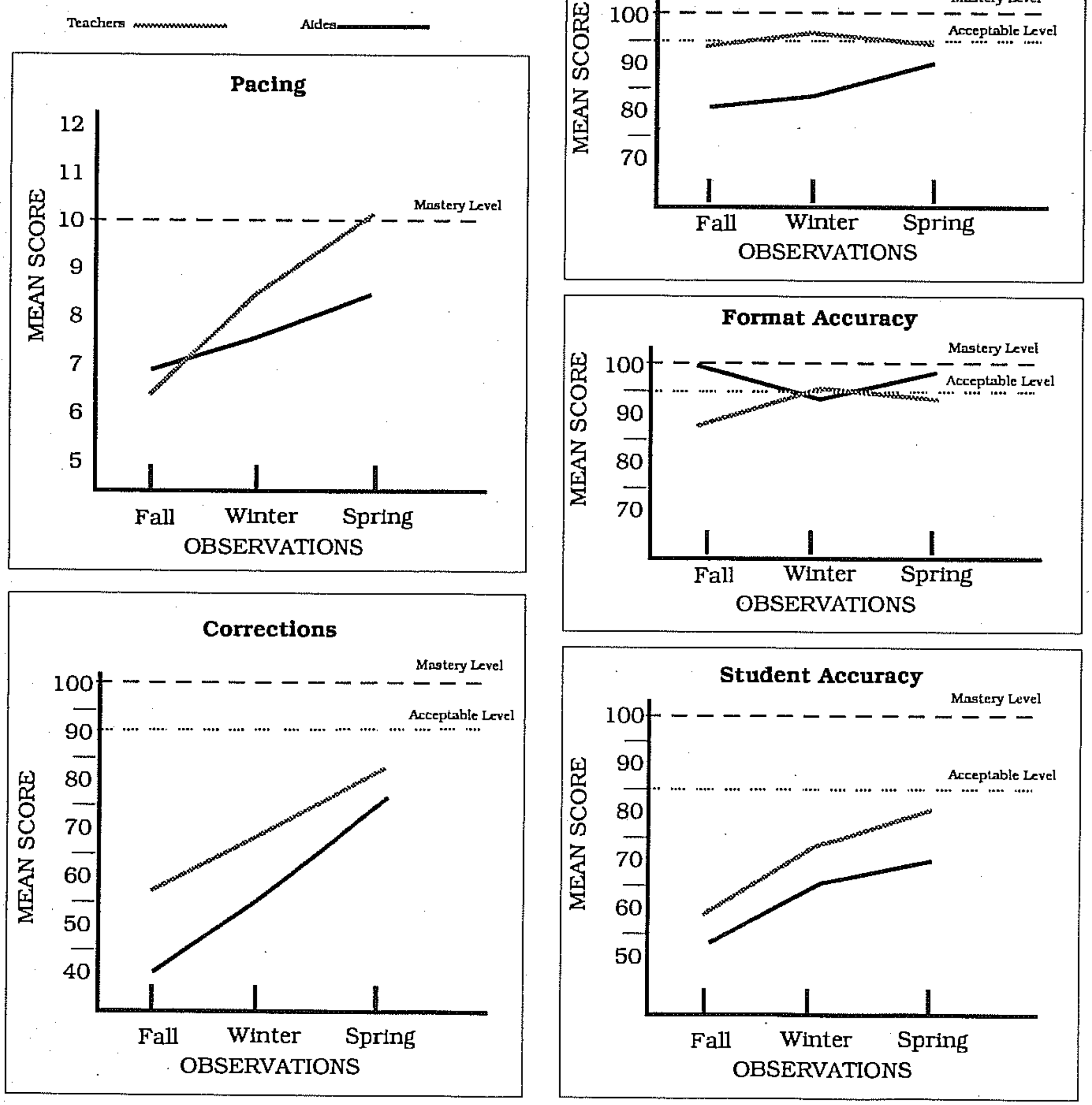
The Classroom Observation Study

During the 1978-79 school year a DI Supervision Code (DISC) was developed to observe the following five critical teacher performance variables: (1) Corrections, (2) Pacing, (3) Format Accuracy, (4) Signals, and (5) Mean Student Accuracy.

Each teacher (or aide) was observed for at least one 12-minute session per phase teaching a small group DISTAR lesson in either reading (for the teachers) or oral language (for the aides). Observers were trained DISTAR consultants. Each teaching trial was recorded in one frame with three components—teacher presentation, student response, and teacher response. Any errors the teaching staff made in presentation (e.g., inaccurate format) or response (failure to correct an error) were noted. If even only one student failed to respond when a unison response was required, the box was marked NR (no response). At the end of 12 minutes, the teacher's pace (number of learning trials per minute) and the accuracy rate of the other four variables (in percentages) could be calculated.

The analysis revealed some reasonably clear patterns of skill acquisition across samples of teachers and paraprofessionals. Two sets of teaching techniques (Format Accuracy and Signals) are acquired by virtually all teachers and most paraprofessionals within a 2-month time span (see Figure 1). On the other hand, many teachers seemed unable to reach criterion level after 8 months on Corrections and Student Accuracy. Acquisition rate on these variables was even slower for paraprofessionals. There were some strong individual differences in acquisition rate. Those teachers ranked in the upper quartile by supervisors (high level imple-

Figure 1. Skill Acquistion Trends for Teachers and Aides



mentors) differed from the lowest quartile only on three: Pacing, Student Accuracy, and use of Corrections. More importantly, there were strong differences not only in observed behavior, but in reading scores on standardized achievement tests between the high implementors and low implementors. (The two high implementer's classes were at the 59th percentile, while the two lowest were at the 27th and 22nd percentile level.)

During the summer and fall of 1979, the observation code (DISC) was streamlined by Adrienne Allen and the authors. Only the three most crucial variables—Pacing, Corrections, and Student Accuracy—were included. Also, an Implementation Rating Form (IRF) was developed by Russell Gersten and Linda Meyer, with extensive input from members of the consultant staff, to offer a broader range assessment of implementation of the DI Model. The IRF rated placement of students, type of corrections employed by teachers, criterion teaching time, and basic classroom management skills. The IRF was completed for all teachers in the project. Correlations between total IRF scores and reading gains were quite strong for both the CTBS and the WRAT (in the range of .54 to .94 with a median of .70).

The study showed that model implementation can be measured in a valid, reliable fashion. The rating form appears to also be a valid instrument, but interrater reliability indicated supervisor bias can be a serious problem in some cases. In contrast, it appears that by observing only three teacher performance variables (Pacing, Corrections, and Student Accuracy) one can predict reading achievement gains for the entire class with a high level of accuracy. There are now two validated instruments available for measurement of DI implementation—the DISC and the IRF.

Case Study Component

Four reports were released by the independent subcontractor discussing administrative aspects of implementation, intensive interviews with teachers and paraprofessionals, and a naturalistic study of the role of the project manager and DI consultants. Only certain key findings will be reported here.

1. *The importance of the consultant and local supervision.* A consistent finding was that most teachers and aides found both the consultants from the University of Oregon and the local supervisors (called resource teachers) extremely helpful, perhaps the most positive aspect of the model. (This was also cited by the court-appointed team.) Most teachers liked the concrete, specific "hands on" type of supervision offered, and the concrete suggestions and feedback offered. Many instructors said they found this model very strange at first, a little awkward and embarrassing. But they came to really appreciate it.

There were some divergent opinions, and some criticisms were also voiced. Many found the expectations held by supervisors during the first year of implementation too high; observations were thus seen as punishing experiences. Several cited that there was a bit of disagreement between the three University of Oregon DI consultants (who each visited once every six weeks or so). Teachers found this very confusing. Some of the local supervisors were considered inadequately trained during the first year. A small, but vocal minority of teachers indicated that they did not care for the entire model of supervision.

Interestingly enough, one of the major findings of the Rand report on implementation of educational innovations was that programs tend to succeed when supervisors offer concrete, specific technical assistance.

2. *Peer support.* Teachers new to the DISTAR program had a much easier time dealing with the complexities of the program when at least some of their fellow teachers were familiar with the model.

3. *Attitudes toward DISTAR.* There was a wide range of attitudes toward DISTAR. One group liked structured programs, liked structured supervision, and seemed to thrive in this program. The largest group, however, was more ambivalent. The found the "mechanical nature" of DI went against the educational values they previously held, and made them feel a little less like teachers. Some also indicated that teaching was now boring (others said they enjoyed not having to spend as much time writing lesson plans, etc.). A common theme in most of this group, though, was that the reading and language gains made by their students were at a level they did not believe possible. So they were willing to continue with the program.

A final group indicated they disliked Direct Instruction. Several wondered if the gains made by the students would last. Some of these teachers were considering transferring to other schools.

Administrative Issues

1. *Start up.* There is a need for a start-up time longer than the one-month period allotted. In the fall of 1978, interested teachers were told to teach DISTAR with only two days of preservice training; the local supervisors had only one week of training and were not equipped to deal with many problems. Furthermore, many administrative decisions (see below) were being made while implementation was ongoing.

2. *Program conflicts.* Conflicts occurred with Title I, Bilingual and Special Education Programs. The "pull out" model used in these programs conflicted with the in-class nature of both instruction and supervision in DI. Conflicts in regulations were not resolved prior to implementation.

3. *Teacher transfers.* It is quite likely that there will be a set of teachers like those in the third group cited above—those who do not like the model for personal or ideological reasons and wish to transfer. Dealing with these transfers is an issue that must be considered.

4. *Role of central administration.* In the Spring of 1979 (the end of the first year), several members of the central staff (including the Deputy Superintendent) made strong statements supporting the DI model. All participants agreed that this had a powerful effect on the participants. However, there was an awareness

that this would mean little without follow-up support from resource teachers and consultants.

5. *Role of building principals.* Virtually every teacher and aide found the building principal *irrelevant* to implementation; she or he seemed to neither support nor hinder the process. (There was one exception in the first year—a principal who actively hindered the work of the consultants.) This would seem to corroborate the findings of the court-appointed team that these principals did not serve as instructional leaders.

Interviews with the 7 principals were conducted by the independent researcher. He reported that most principals regretted their inability to deal with the instructional sphere; 3 of the 7 asserted that they perceive this as their key role. However, all agreed that, due to the presence of at least 20 disparate budgets at their schools, and the multitude of Federal programs, they were forced to operate as managers.

Future Research

Research being conducted this year [1981] for the National Institute of Education attempts to look more seriously at the role of the principal in managing implementation and in making sure teachers receive adequate technical assistance. Both ethnographic and quantitative methods will be employed by: (1) looking at the current situation at target schools vis-a-vis technical assistance, monitoring of teaching and student progress, instructional leadership issues, (2) providing a series of inservice activities to interested principals based on prior needs assessments. This will include increasing knowledge about effective teaching, use of criterion referenced tests to monitor progress, and basic principles of supervision, and (3) evaluating changes in the school system—especially in terms of amount of time students spend actively engaged in learning. ◆

A Study of Teacher Presentation Variables— Pacing and Praise

by Craig Darch, Auburn University
Russell Gersten, University of Oregon

From *ADI News* Volume 2, Number 4 (Summer, 1983)

Identification of instructional presentation variables that lead to increased student performance is an important aim of educational research. Presentation variables are especially important in programs for mildly handicapped students. Two procedures essential to success for these students are rate of presentation and use of praise. Both promote attention to task and accurate work. Success in these areas is important in the development of academic achievement and self concept.

Rapid pacing has been shown to improve the academic performance of non-handicapped students (Carnine, 1976). Many studies show that teacher praise is important in all aspects of an educational program (e.g., Walker, 1979).

The present study was designed to extend the research on rapid pacing and feedback to learning disables students. Also, both the isolated and interactive effects of rapid pacing and praise were evaluated.

Method

Subjects and Setting

The subjects for this study were four students, three boys and one girl, who were identified as learning disabled by school psychologists. Each had a history of difficulty in basic word decoding. At the beginning of this study, each subject had just been placed into the learning disabilities classroom. Students' individualized educational programs (IEP's) included learning basic sound/symbol relationships and blending sounds into words.

The reading class met daily for a 45 minute period. The first 20 minutes were used for the experimental intervention. The remainder of the class period was spent having the students work on individually designed worksheets. The four students were taught as a group. The teacher for this study was a 23-year old woman who was completing a graduate practicum in learning disabilities.

Reading Program

The DISTAR Reading I Program was used for this study. In this program, the rate of teacher presentation (pacing) is expected to be rapid (approximately 10 tasks per minute).

Before the study began, the teacher and experimenter practiced providing signals for student responding, contingent praise, and increasing instruction. Training consisted of demonstration, practice, and feedback. These sessions occurred after school and lasted approximately 30 minutes for five consecutive days.

Experimental Design

A modified reversal design, with replication across subjects, was used. The advantage of this particular design is that it allows demonstration of both the isolated and interactive effects of individual intervention components. Because most instructional systems are multi-component, the ability to identify those aspects of a treatment that contribute most to increased student performance is very important. Data were collected daily during this study and experimentation lasted a total of 25 school days.

Baseline 1

During this phase, the teacher presented material from the DISTAR lessons by closely following the daily lesson scripts, with two exceptions. Although the teacher told students whether their responses were correct or incorrect, she did not verbally praise correct answers or instances of appropriate social behavior. If the group gave a correct response, the teacher would merely tell the students the answer was correct, and move on to the next step in the lesson.

The rate of teacher presentation of the instructional content was also controlled. To insure a slowly paced instructional delivery, the teacher was instructed to pause five seconds (by counting to herself) between the end of a student's response and the presentation of the next part of instruction. This procedure was similar to the one used by Carnine (1976). Baseline 1 lasted four days.

Rapid Pace 1

This phase lasted four days and was exactly the same as Baseline 1 except that the pace of instructional presentation was increased. Instead of creating a five-second pause between the completion on one task and the introduction of the next, the teacher was instructed to proceed immediately to the next part of the lesson. It must be emphasized that increased pacing had nothing to do with the rate at which the teacher spoke. Instead, pacing was defined as the time between task presentations.

During this phase, the teacher gave only informational feedback to the student; she did not verbally praise their behavior. Corrections were done in the same manner as in Baseline 1.

Rapid Pace and Praise 1

This phase replicated the rapid pacing procedure and added praise for appropriate social behavior and correct academic responses. During each instructional presentation, the teacher identified instances of appropriate behavior and praised the students, using short specific statements. The teacher reinforced behaviors such as eyes focused on the text, keeping hands and feet to oneself, staying in one's seat, and responding to the teacher' signal. Also, the teacher would praise student(s) or the group for responding correctly. This phase lasted five days.

Praise Only

Next, rapid pacing was dropped to demonstrate the differential effectiveness of praise vs. pacing. The teacher again presented material at a slow pace, using the five-second-pause technique. The teacher replicated the verbal reinforcement procedure described in the previous condition. This phase lasted three days.

Baseline 2

This phase was identical to Baseline 1 and lasted three days.

Rapid Pace 2

This phase replicated Rapid Pace 1 and lasted three days.

Rapid Pace and Praise 2

This phase, a replication of Rapid Pace and Praise 1, was important again showing the effects of the two variables in combination. It lasted three days.

Dependent Variables

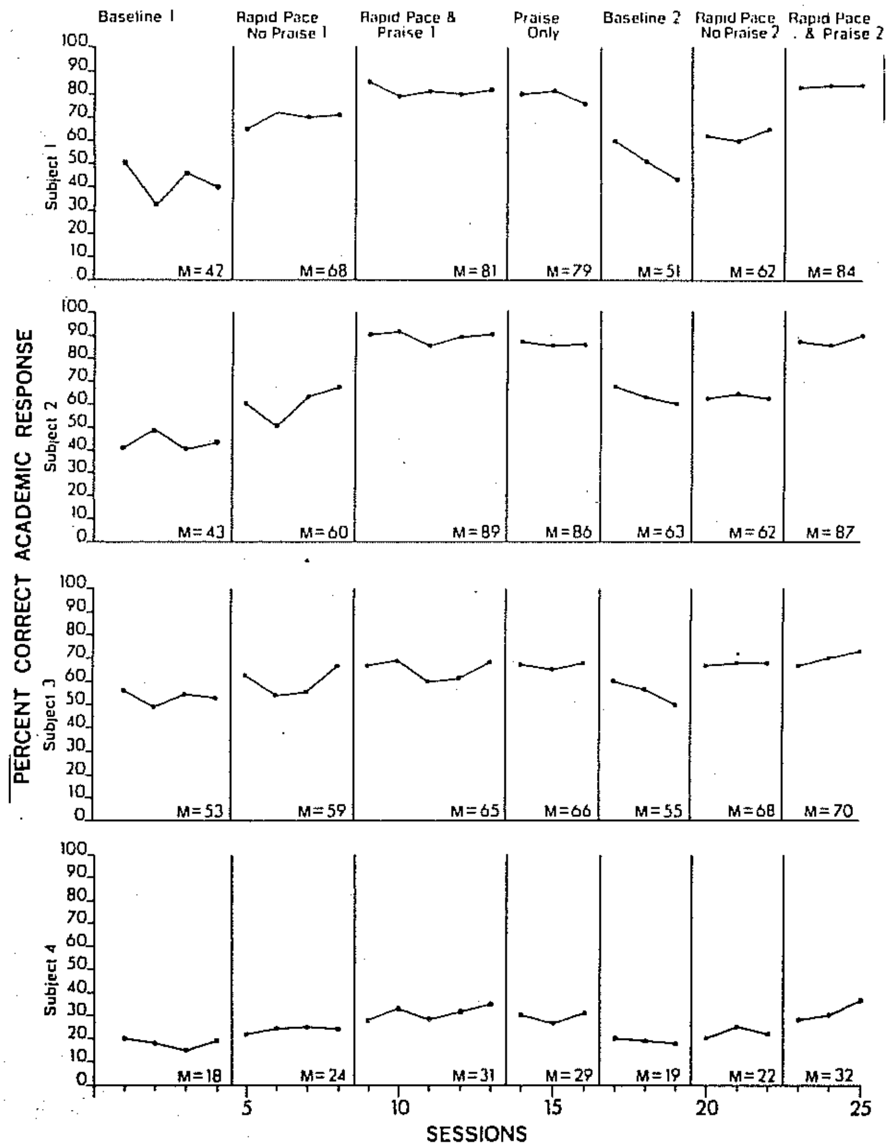
Percent Correct. Percent Correct was calculated by dividing the number of opportunities available to respond to a teacher question or direction by the number of correct answers. If the student did not correctly respond within one second of the teacher's signal, the response was marked as incorrect. In order for a student's response to be considered correct, the student had to voice the answer loudly enough to be heard by the observer, who was seated approximately four feet away. Data were taken during both individual turns and group responses. Procedures are discussed below.

Percent On-Task Behavior. On-Task was defined as eyes on the teacher or the presentation book, keeping hands and feet to oneself, and responding on time to the signal. Behaviors such as talking out, talking to classmates, playing with objects, out-of-seat without permission, or fighting were classified as off-task.

Data Collection Procedure

Three university graduate students served as primary data collectors. One collected data on On-Task Behavior while the other recorded Correct Responses. On alternating days, the third graduate student would observe and record either On-Task Behavior or Correct Responses to assess inter-observer reliability. Data were collected on the four subjects in the following manner. Observers who coded On-Task behavior used a ten-second-time-sampling technique; they would observe the student for nine seconds and record the student's behavior for the entire nine seconds, the observers would record a slash (/) in the appropriate interval on the data sheet. If the student was engaged in any inappropriate behavior during the nine seconds, the observers recorded a 0 in the specific interval. A different student was coded each ten-second interval. The observers systematically coded the four subjects in a predetermined order; therefore both observers were recording the same subject at the end of each ten-second interval. One complete rotation through the subjects took forty seconds and the process was repeated continuously for the duration of the session.

Figure 1.



The technique to record Correct Responding followed the procedures described by Carnine (1976). Each subject was recorded individually for five consecutive tasks, *even though all subjects responded to each task.* A task was defined as a teacher presentation of either a sound or a word. For example, each task required the students to: (a) give the sound that a letter or letter combination made, (b) sound out a word, (c) to read the word as a whole unit. This cycle was repeated numerous times during the 20 minute lesson.

Results and Discussion

Inter-observer agreements for on-task behavior ranged from 80% to 95%, with a mean of 90%; for percentages of correct answering, they ranged from 88% to 100% with a mean of 95%.

Accuracy

The results on the percentage of Correct Responses for each subject are shown in Figure 1. Each student was performing well below acceptable level during the Baseline 1 condition. Each student's performance was affected by the introduction of each phase of the experimental intervention. With the introduction of the first component of the intervention, Rapid Pace 1, each student demonstrated an increase in *accuracy*. This increase in performance was replicated in the Rapid Pace 2 condition. This consistent effect shows that teachers of LD students can improve academic performance merely by increasing the rate of instructional presentation. This finding replicates Carnine's (1976) study. This result also supports the findings of a correlational study conducted by Gersten, Carnine, and Williams (1982), in which the students of teachers who consistently paced their lessons briskly gained more in reading achievement than their peers. It is important to note that this increase in pacing was not due to the teacher's speaking more quickly. Rather, the pauses between each segment of the lesson were reduced.

The combination of Rapid Pacing and Praise led to an even stronger effect. Students 1 and 2 improved from a baseline accuracy of 40% to 86% and 79%, respectively. Student 3's growth was not as dramatic (from 53% to 65%). Student 4 was the lowest performer in the group. Though he demonstrated some growth in accuracy (18% to 31%), he continued to perform well below acceptable levels.

When the praise only condition was administered, performance gains on Correct Responses are maintained in student 3, while slight decreases are shown for the three other students.

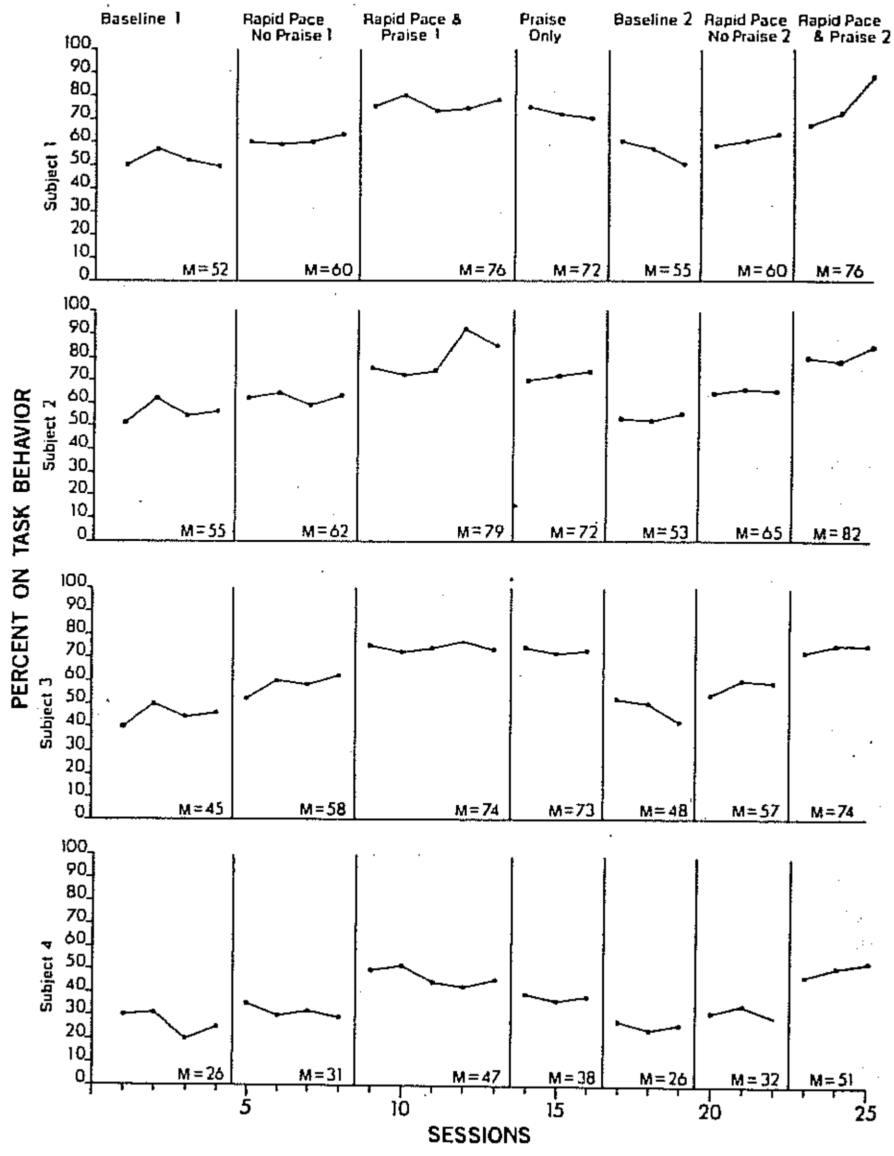
Levels of performance during the last three phases replicate the results in the first three conditions of the experiment. The finding that the combination of Rapid Pacing and Praise can produce powerful increases in performance of LD students is an important finding for the classroom teacher.

On-Task

The results for On-Task Behavior (Figure 2) replicate those for Correct Responding. Rapid Pacing and Praise both contributed to increased On-Task Behavior. These findings have implications for teachers whose students have high levels of Off-Task Behavior.

Teachers must look to both instructional modifications and use of consequences when developing instructional programs for skill deficient students. The powerful effects of pacing and praising can be important in allowing students to succeed in the early stages of a remedial education program. The finding that

Figure 2.



rapid pacing alone increases performance might be helpful as a method to eliminate (or reduce), at some point in a student's instructional program, supplemental reinforcement programs (i.e., token systems, high rates of praise). A number of investigators have suggested that use of extrinsic reward systems may produce students who only learn to earn tokens. These students may fail to develop intrinsic motivation for academic tasks. If teachers modify instructional delivery variables (i.e., pacing) some students may not need supplemental reinforcement programs. Further, if a student has already been placed on a point system, an increased instructional pace may facilitate early quick removal of this intervention. ◆

References

Carnine, D.W. Effects of two teacher presentation rates on off-task behavior answering correctly, and participation. *Journal of Applied Analysis*, 1976, 9, 199-206.

Gersten, R.M., Carnine, D.W., and Williams, P.W. Measuring implementation of a structured educational model in an urban school district: An observational approach. *Educational Evaluation and Policy Analysis*, 1982, 4, 67-79.

Walker, H.M. *The acting out child: Coping with classroom disruption.* Boston: Allyn and Bacon, 1979.

Focus on Student Performance– *The Key to Effective Supervision*

by Mary Gleason
University of Oregon

From *ADI News* Volume 3, Number 4 (Summer, 1984)

A premise of the Direct Instruction Model is that all children can be taught if they are provided with adequate instruction. The role of the supervisor is to help the teacher provide adequate instruction, so it follows that if the supervisor helps the teacher, the supervisor has helped the children to be taught. The measuring stick of the teacher's success, and of the supervisor's, is the academic success of the children. Supervisors must monitor teacher performance, and their own, by monitoring student performance.

Many supervisors and administrators feel that they must approach a teacher's classroom armed with data forms. Data forms tend to be written only in terms of teacher behaviors, not in terms of child performance. Some supervisors get sidetracked. The ultimate focus of the supervisor's observation should be student learning. In monitoring student learning as well as teacher performance, data forms are useful tools, but should not be the supervisor's only tool.

In classrooms where Direct Instruction programs are being taught, the supervisor or administrator has two expectations: (1) students will cover a lesson a day in each Direct Instruction program, and (2) students will perform at a high success level. These two expectations represent the outcomes the supervisor is looking for. All observations in the classroom are ultimately concerned with whether these two expectations are being met. (These statements do not deny that we also want the children to be having fun and to feel good about learning.)

If the observations yield the information that children are learning and at an acceptable rate, the supervisor has reason to reinforce the teacher. If, on the other hand, the children are not being taught as well as they could be, the supervisor offers practical suggestions for change. Effective teacher change equals improvement in student performance.

What a Supervisor Looks *For*

Time allocated. First, the supervisor should look at the teaching schedule to make sure that enough time has been allocated to be able to do a lesson a day. Children will not complete *DISTAR Reading I* in one year if the teacher allows 20 minutes a day for the program. If a particular group of children can't get firm on a lesson in one day, the teacher may have to schedule another period of teaching time for that group.

Lessons covered. After the supervisor has checked the teaching schedule, he/she should help the teacher design a way to keep track of how many lessons are being covered. One way is to keep track of the lesson gain of each group on a weekly basis. For each group, the teacher would write down the number of the lesson worked on that day. At the end of the week, the teacher would write in the total number of lessons covered that week (see Figure 1).

Figure 1.

Week 1					
Mon	Tues	Wed	Thur	Fri	Gain
53	54	55	55	56	4

Appropriate placement. The supervisor should check for appropriate placement of the group. The children should always be performing at a high enough success level that they can feel good about working hard. When children are "over their heads," they have difficulty staying on task and the teacher spends too much time correcting and firming.

The supervisor can check for a high success level in a number of ways: (1) by looking at the results of a criterion-referenced test for each child in the group to see if each child is performing between 80% and 100%, or (2) by taking data on students' oral responses during instruction, looking for 80% or higher on first-time responses (correct responses after a correction don't count) and checking students' independent work performance, looking for 80% or higher on worksheets, and 97% or higher on oral reading.

The supervisor can continue to use these procedures to monitor whether students are being moved on to new lessons before they have mastered the material. When the students are "over their heads," they should be moved back to a lesson where they can be more successful.

Physical arrangement. The supervisor should check for physical arrangements, organization of materials, and use of time that enhance the teacher's ability to teach well. Are all children seated so they can see the teacher and the material used for presentation? Are the lowest performers sitting closest to the teacher? Are the teacher's materials close by and organized so that no time is wasted in transition from task to task?

Frequent responses. The supervisor should look to see if the teacher is getting frequent responses from the children. The supervisor can check response rate (pacing) by doing the following: During a five-minute period, make one tally point each time the students respond orally. Divide the number of tallies by 5. A response rate of 2 to 7 responses per minute means the teacher is talking too much going, too slowly, or is somehow wasting time. Approximately 10 responses per minute indicated an effective response rate.

Student errors. The supervisor should watch the children. He/she should pay attention to student errors and what the teacher does to "firm" the children's skills. It is possible for a Direct Instruction teacher to "look" technically perfect and still have children who are not firm. The teacher's pacing is great, the signals are precise, and, every time an error is made, the teacher does a correction; however, the teacher allows the children to move on to the next lesson while they are working at a 60% success level.

This type of teacher can fool an unsuspecting supervisor who watches the teacher's presentation and forgets to attend to the children's performance. This teacher needs as much help as the teacher who has poor signals. The supervisor should watch for the following: Does the teacher stop at each error and immediately tell the answer? After telling the answer, does the teacher repeat the missed task so the children can try again? Does the teacher go on to something else and then come back to the missed task to see if the students can perform correctly following a delay? Does the teacher repeat the format that students made errors on before going to the next format? Does the teacher check all written work and provide a correction for each item that is missed?

Student errors also occur because of the teacher's presentation skills. When a supervisor sees student errors, the supervisor must try to determine of the errors are caused by poor signals, inappropriate thinking time, or other teacher behaviors.

On-task behavior. The supervisor should check whether all students are working all the time and whether the teacher takes steps to teach students to attend and work hard. When the teacher is asking for unison responses, the supervisor must watch to see if all students are answering and if they are answering together.

For those who are just beginning to use the Direct Instruction model of supervision, perhaps you can get started by using a simple checklist (see Figure 2). And remember, keep your eye on the kids. ◆

Figure 2.
What to look for in a Direct Instruction classroom.

1. Time allocation for each group.
2. Amount of content covered.
3. Appropriate placement.
4. Physical arrangement, organization of materials.
5. Smooth, rapid transitions
6. Frequent responses.
7. Student errors
8. On-task behavior.

Barriers to Educational Change

by Douglas Carnine

From *ADI News* Volume 1, Number 4 (Summer, 1982)

> *"The major studies of curriculum reform have shown that where training, the introduction of materials, vertical political solidarity, and staff and administrative commitment are brought together, there is considerable movement. Gradually, however, the school returns to the normative patterns which characterize most American schools and the innovations lose their steam. The problem is a worldwide one."*
>
> (Hersh, Carnine, Gall, Stockard, Carmack & Gannon, 1981)

Even though teacher's greatest rewards have to do with serving their students (Dunn, 1980), innovative practices that help teachers better serve students are under-utilized or misused. The defenses operating to resist change may be labelled *discrediting, delaying, distorting,* and ultimately, *discontinuing.* We will discuss each of these, drawing in part from our fourteen years of experiences with Project Follow Through. The Follow Through experiment has provided clear evidence that Direct Instruction methods can be effective in overcoming educational problems of the economically disadvantaged (Becker & Carnine, 1980; Stebbins, St. Pierre, Proper, Anderson & Cerva, 1976, 1977), and yet many choose not to believe this evidence.

Discrediting

If an innovation is discredited, pressures to adopt it are minimal. Innovations are usually discredited through intellectual or quasi-intellectual activities such as attributing their success to unique factors not found in other settings, questioning the values represented by the innovation, questioning, criticizing and ignoring any evaluation that judges the innovation to be effective, or even claiming that the innovative practice has already been adopted when, in fact, it has not.

According to the uniqueness argument, the effective instructional program that operated for over a decade at P.S. 137 in the Ocean-Hill Brownsville section of new York could not work in P.S. 73, which is located only three blocks away — despite the fact that the program in P.S. 137 operated effectively for 12 years with 5 different principals, 4 different compensatory eduction directors and over 50 different teachers. If this view is true, scientific work in education is a contradicition in terms, since, to be valid, scientific knowledge must be true in more than one setting. What would be the reaction to a doctor who said that heart surgery could be done successfully on 14th St. in Chicago, but not on 11th Street? In education, people readily accept data showing that most students can read in one urban school while most students cannot read in another school a few blocks away, without ever considering that this may be due to the specific instructional processes going on in the two buildings. Early sociological data (Jencks and other, 1972; Coleman, Campbell, Hobson, McPartland, Mood, Weinfield & York, 1966) has been used to support the belief that schooling makes little difference, although more recently, even Jencks seems to be attributing more importance to the schooling process (Jenks, 1979).

Undermining an innovation by questioning the values it represents is more subtle. "For the educationists, the doctrine of the whole child is the magical balm that washes away their sins. Ask a question about skills, and you get T.S. Elloit, transforming the question to one about values" (Lyons, 1980). The Direct Instruction Model, which has been relatively effective in fostering both academic growth and a positive self-concept in economically disadvantaged children was seemingly discredited on a PBS television documentary by a survey of principals who said that the program does not address creativity and other aspects of the whole child, and that while the programs' effectiveness was well known, it was believed to be insufficiently "humanistic" turning teachers and children into robots. One principal new to a building forbade two teachers to continue the program even though he had never seen it in use and even refused to observe teachers using it.

Similarly, critics can claim that any evaluation of an innovative program is invalid because it doesn't measure what is truly important. As Anderson and colleagues state, ". . .Any program that wishes to rid itself forever of the discomforts of evaluation need only add to its list of objectives one metaphysical, obscure, or otherwise immeasurable purpose. . ." (Anderson, St. Pierre, Proper & Stebbins, 1978). Cognizance of both intended and unintended effects of an innovation is important. It is always necessary to limit the generalizability of an evaluation effort to what it measures. However, to reject an evaluation because it did not measure what may be impossible to measure (e.g., the inner feelings, aspirations, or creative potential of a generation of school children) is unreasonable. Nonetheless, an unpopular evaluation finding can be discounted by emphasizing what was not measured.

Another way of discrediting a successful innovation, one that may seem almost inconceivable to the naive reader, is to simply ignore its success. "Although pupil achievement data are routinely collected for individual students and are used to monitor their progress and determine their opportunities, the same data are rarely aggregated so as to provide a basis for assessing the performance of individual teachers, schools, or districts" (Meyer, Scott & Deal, 1979). By failing to aggregate data and compare progress across schools and classrooms, administrators relieve themselves of the responsibility to either provide remedies to low performing schools or explain why some schools are effective.

Delaying

Even if an innovation is not discredited, its adoption can be delayed. While delaying is characteristic of many fields, there are some exceptions, such as medicine, in which technological advances are often rapidly accepted. In one study, a miracle drug was adopted by 90% of the physicians in four communities within 7 months. Typically, the complete adoption of an educational innovation is at least ten times slower (Carlson, 1964). The medical comparison is probably unfair. Many past educational innovations have been fads that proved to be of little benefit. Thus, slow-paced adoptions have served to minimize useless disruptions to school systems. As valid practices become more prevalent in education, delay may become less of a problem, but that seems unlikely.

Distorting

No innovation is implementation-proof. Innovations that are not discredited and delayed can still turn out to be ineffectual as a result of extensive modifications. For an untried innovation, a process of adaptation seems reasonable; in fact, all things being equal, the more an innovation is adapted, the more likely its acceptance in a school (Berman & McLaughlin, 1975). Too often, though, adaptation becomes a euphemism for distortion. For example, Centra and Potter (1980) cited several studies of "team teaching" in which the investigators could not even identify which teachers were working in teams. The innovation had been transformed in such a way that it no longer differed from traditional practice. Another common occurrence is selecting only a part of the innovation for implementation (in the name of eclecticism) and then attributing the subsequent failure to the entire innovation. Finally, an innovative practice adopted by a district may never be implemented, because of what Charters and Jones (1973) refer to as the "risk of appraising non-events."

Sometimes distortion seems more like sabotage. For example, in one small urban school, an innovative program was adopted that required para-professionals and specific instructional materials. One year the central administration delayed hiring paraprofessionals for eight weeks, even though trained people and funds were available; over 3200 hours of instructional time were lost. A few years earlier, the same district delayed an order for essential instructional material for over six months, resulting in the loss of thousands of hours of instruction. Disruptions can also be effected by transferring key personnel to other schools or bringing in personnel opposed to the innovation.

Discontinuing

Innovative practices, even those that are eventually implemented and proven to be quite effective, are often discontinued. Rowan (1977) found that innovations that had the least to do with instruction (e.g., school health and cafeteria services) had the greatest likelihood of survival. Those indirectly related to instruction (such as guidance counseling, and psychological testing services) had a moderate likelihood of survival. However, innovations that actually dealt with instruction were the least stable, and were terminated most quickly.

Abandonment of a program by administrators can even occur in the face of public support. At a school board meeting for a small rural community, several parents testified in support of a relatively new, highly structured compensatory education program. One parent's three oldest children, who started school before the district installed the new program, hadn't learned to read. Later, two younger children, who had the benefit of the new program, tutored their older siblings. The parent was worried that her sixth child, only four years old, would be a school failure, too, if the program were dropped. Despite the district's acknowledgement that the program was quite effective with poor children, the board voted to discontinue it after teachers charged that the program was too structured and too narrow in outlook.

A Plan of Action

Reviewing case studies and theories of change as well as conducting assessment can help policy makers identify variables crucial to the success of an innovation and predict the conflicts that might occur during the change process. Based on that information, planners can then develop enhancement strategies to alter or work around anticipated barriers. These strategies are usually based on both authority and consensus (Greenwood, Mann & McLaughlin, 1975). Suppose most of the middle-level managers in a district (principals, curriculum specialists, teacher trainers) oppose the impending implementation of a major innovative practice. Superintendents might exercise their authority by visiting the project and making statements about its importance. For consensus–building purposes, the middle-level managers might be paid to attend an out-of-town training session, attended by enthusiastic users of the innovation from other school inside and outside the district. A combination of practical, common sense advice from session leaders, coupled with genuine testimonials from peers, could contribute to a willingness to give the innovation a chance.

Diplomatic negotiation with hostile middle-level managers is one possible way to forestall an innovation from being discredited. The general strategy is to anticipate how the innovation might be discredited and focus on those points — have potential adopters observe the innovation in action, talk to current users, review pertinent research reports and papers by popular "opinion leaders," and establish training programs in the innovative practice.

A different strategy is necessary in responding to delay. Situational leadership theory suggests that if an administrator is not interested or energetic about change, others must become task oriented and assume responsibility for planning and interaction (Hersey & Blanchard, 1977). As interest in the innovation grows, responsibility for the implementation can be shared even more. The work of Tannenbaum and Schmidt (1973) on leadership styles (selling, telling, consulting, testing, joining, delegating) is relevant to the process of shifting responsibility to those who will carry out the innovation. Sometimes distinguishing between reasonable, inevitable slowdowns and destructive delays is difficult. Once delays clearly begin to undermine the innovation, however, pressure should be applied. The push can come from above or from peers. Either way, a significant blockage must be removed or satisfactory implementation may never come about.

In deciding what constitutes a significant blockage, it is important to identify the critical, non-negotiable aspects of the innovation. Those aspects must be kept clearly in mind during installation. Otherwise, a harmless adaptation cannot be distinguished from a major dilution of the innovation. Since adaptations should be encouraged and distortions discouraged, the distinction is critical. Without protection from significant distortions, an innovation will quite likely fail. Conversely, fighting inconsequential adaptations can wear out everyone and create animosities among people who need to work together.

Knowing when and how to fight deviations is only part of the strategy for dealing with distortions. Another critical component is supporting genuine attempts to implement the innovation. Gersten and Carnine (1981) have identified several support tasks, culled from research on effective school and classroom practices. Some of these tasks are: (a) assessing how well the innovation is being implemented in each classroom, (b) procuring appropriate technical assistance, and (c) directing rewards and sanctions according to the quality of implementation. Since most principals are unlikely to carry out these tasks, responsibility for them must be shared or delegated. The principal might work in coordination with a school-level supervisor, possibly a lead teacher.

With concentrated thought and effort, practices based on effectiveness research can become institutionalized. Schools would then assume greater responsibility for utilizing effective practices while still fulfilling institutional requirements. ◆

References

Anderson, R.; St. Pierre, R; Proper, E.; and Stebbins, L. "Pardon Us But What Was the Question Again?" *Harvard Educational Review*, 48 (1978): 161-70.

Becker, W.C., & Carnine, D.W. Direct instruction: An effective approach to educational intervention with the disadvantaged and low performers. In B. Lahey and A. Kazdin (Eds.), *Advances in Clinical Child Psychology*, Vol. 3. Plenum Publishing Corp., 1980.

Berman, P., and McLaughlin, M.W. *Federal Programs Supporting Educational Change. Vol. IV: The Findings in Review.* Santa Monica, Calif.: Rand Corporation, 1975.

Carlson, R.O. "Environmental Constraints and Organizational Consequences: The Public School and its Clients." *NSEE Yearbook*, University of Chicago Press, 1964.

Centra, J.A., and Potter, D.A. "School and Teacher Effects: An Interrealational Model." *Review of Educational Research*, 50 (1980): 274-91.

Charters, W.W., and Jones, J. "On the Risk of Appraising Non-Events in Program Evaluation." *Educational Researcher*, 2 (1973).

Coleman, J.S.; Campbell, E.Q.; Hobson, C.J.; McPartland, J; Mood, A.M.; Weinfield, F.D.; and York, R.C. *Equality of Educational Opportunity*. Washington, D.C.: U.S. Government Printing Office, 1966.

Dunn, P. "Motivating Teachers for Excellence." Fact Sheet, Number One. Eugene: ERIC Clearinghouse on Educational Management, University of Oregon, 1980.

Gersten, R., and Carnine, D. *Direct Instruction Mathematics: A Longitudinal Analysis for 2,000 Children. Technical Report 81-1. Eugene, Ore.*: University of Oregon, 1981.

Greenwood, P.W.; Mann, D.; and McLaughlin, M.W. *Federal Programs Supporting Educational Change, Vol. 3: The Process of Change.* Santa Monica, Calif.: Rand Corporation, 1975.

Hersey, P., and Blanchard, K.H. *Management of Organizational Behavior: Utilizing Human Resources, 3rd ed.* Engelwood Cliffs, NJ: Prentice-Hall, 1977.

Hersh, R.H.; Carnine, D.; Gall, M.; Stockard, J.; Carmack, M.A.; and Gannon, P. "The Educational Professions and the Enhancement of Classroom Productivity." Eugene, Ore.: Center for Educational Policy and Management, University of Oregon, 1981.

Jencks, C. *Who Gets Ahead: The Determinants of Economic Success in America.* New York: Basic Books, 1979.

Jencks, C., and others. *Inequality: A Reassessment of the Effect of Family and Schooling in America.* New York: Basic Books, 1972.

Lyons, G. "Why Teachers Can't Teach." *Phi Delta Kappan*, 2 (1980): 109-12.

Meyer, J.W.; Scott, W.R.; and Deal, T.E. "Institutional and Technical Sources of Organizational Structure Explaining the Structure of Educational Organizations." Paper prepared for a conference on Human Service Organizations, 1979 at the Center for Advanced Study in the Behavior Sciences, Stanford, Calif.

Rowan, B. "Bureaucratization in the Institutional Environment: The Case of California Public Schools, 1930-1970." In *The Structure of Educational Systems: Explorations in the Theory of Loosely-Coupled Organizations*, edited by M.R. Davis, T.E. Deal, J.W. Merey, B. Rowan, W.R. Scott, and E.A. Stackhouse. Stanford, Calif.: Stanford Center for Research and Development in Teaching, 1977.

Stebbins, L.; St. Pierre, R.; Proper, E.; Anderson, R.; and Cerva, T. *Education as Experimentation: A Planned Variation Model. Vols. 4A-D.* Cambridge, Mass.: Abt Associates, 1977

Tannenbaum, R., and Schmidt, W.H. "How to Choose a Leadership Patter." *Harvard Business Review*, 51 (1973).

Corrective Spelling Program Evaluated

by Mike Vreeland
Kalamazoo Public Schools

From *ADI News* Volume 1, Number 2 (Winter, 1982)

Three teachers and approximately 60 students took part in an evaluation of three spelling programs in the Kalamazoo, Michigan Public Schools during the 1980-81 school year. One of these programs was *Corrective Spelling Through Morphographs*. It will be referred to as Program A. The other two programs will simply be referred to as Programs B and C.

All of the students involved in the study were 4th graders enrolled in regular (non-special education) classes. A district consultant described the three groups as equivalent in reading level at the beginning of the study.

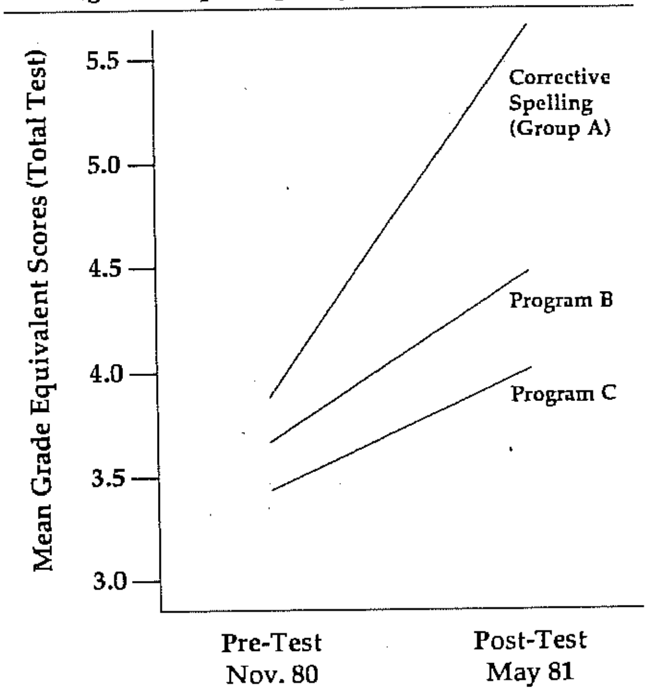
Each of the three teachers taught a group of 20 students for one-half hour each school day for seven months, using one of the three spelling programs studied. Teacher A taught from Program A. Teacher B from Program B, and Teacher C from Program C. Teacher A, who taught from *Corrective Spelling Through Morphographs*, received approximately six hours of training in this program and was observed while teaching and given feedback weekly during the first month of the study and biweekly for the remainder of the project. The other two teachers received no special training in their programs and were not observed or given feedback during ths study.

Students were tested in November and again in May using the *Test of Written Spelling*, a group administered test requiring written answers. The test is divided into two parts—one dealing with phonetically regular words and the other with irregular words. Results of this testing are shown in Figure 1.

Students in Program A (Corrective Spelling) increased from 3.9 to 5.6 in total test grade equivalent scores, an increase of 1.7 grade levels during the seven months of instruction provided in this study. Group B students increased .8 grades from 3.8 to 4.6. In Group C, students climbed from 3.5 to 4.2—a .7 grade level increase. Increases on component parts of the test parallel total test scores. Group A students showed 2.2 and 1.5 grade equivalent increases on the regular and irregular word sections, respectively; Group B improved .6 and .9 on these sub-tests; and Group C's component gains were .8 and .6.

The gains made by the *Corrective Spelling* group (Group A) appear to be clearly superior to those of either Group B or Group C. In both cases, Group A increases were more than twice those of the comparison groups. However, the effects should be interpreted conservatively for a least two increases: (1) instructional time was only estimated to be equal for all groups - this was not measured; and (2) Teachers B and C did not receive program-specific training or supervision, as did Teacher A. This second limitation should be considered carefully, since it could have contributed to the differing program effects as much as the program differences. What we do know, however, is that *Corrective Spelling*, supplemented as it was in the study by teacher training and supervision, can produce gains averaging approximately two-and-one-half months of achievement for each months of instruction. The factors contributing to such instructional efficiency clearly seem worthy of further investigation. ◆

Figure 1. Spelling Program Comparisons.



A Study of 4th-6th Grade Basal Reading Series*

by Ziggy Engelmann

From *ADI News* Volume 1, Number 3 (Spring, 1982)

As part of the development of the Direct Instruction reading programs for grades 4, 5, and 6, we did a rather elaborate study to gain more precise information about teacher behavior and how teacher behavior relates to "the ideal." The design of the experiment was basically simple. We first analyzed the major basal reading programs that are used in grades 4 through 6 - Ginn, Scott Foresman, Houghton Mifflin, and Holt. We we analyzed the programs, we considered the clarity of the communication that was provided, the adequacy of the practice, and other aspects that should be controlled by an effective program. Next, we interviewed the 17 teachers who participated in the study. We provided them with no information about the nature of the study. (They knew only that they would receive some free materials for participating.) Their participation involved answering questions during two taped interviews and video taping two lessons in their reading program. They were told that they would be taped teaching a main-idea lesson and another lesson (whatever lesson came up during the time scheduled for the testing). The teachers were selected from various regions of the United States, from Bridgeport, Connecticut, to Eugene, Oregon.

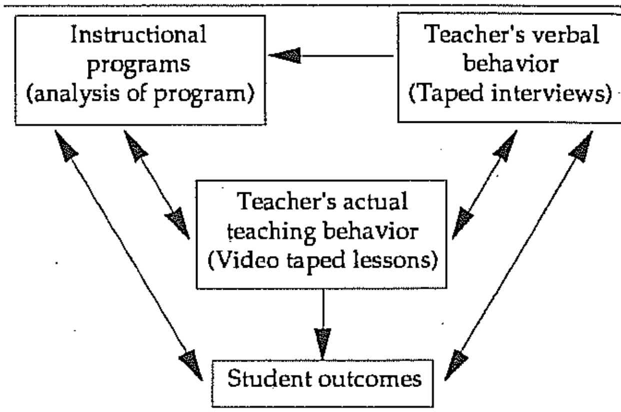
After the students received a lesson, they received a simple test of the material that the teacher had just covered. There were no trick items, no extensions of concepts, and basically nothing more than what the teacher had just taught. The test was sufficiently long to provide a reasonable sample of each student's understanding.

With the information from the analysis of the program, the teacher interview, the record of the teacher's teaching, and the student outcomes, we had the information need to perform a rather thorough analysis that we felt would answer the following questions:

1. Based strictly on an analysis of the program material, how well would the program be predicted to teach the average student?
2. How much do teachers actually deviated from the specifications of the basal programs, and if they do deviate, to what extent do these deviations facilitate communication? (In other words, how much better do the teachers teach than they would if they followed the program to the last detail?)
3. How do the facts about the teacher's instructional program, the teacher's actual teaching behavior, and the actual student outcomes relate to the teacher's verbal descriptions of these areas? (Are teachers accurate and knowledgeable about the details of their programs? Do they know specifically the types of problems their children have? Do they accurately evaluate their own teaching?)

Figure 1 shows the four areas that were investigated. The arrows indicate the various comparisons that were possible from one area to another areas.

Figure 1. Four Areas of Invesitation



Program Analysis

Perhaps the greatest new contribution the study provided was a basis for analyzing instructional programs. The analysis was based on fairly reliable information that we had received when developing Direct Instruction programs. Tryouts consistently disclosed that skills must be taught for a minimum amount of time, that the wording should be simple and consistent, that the skill must be reviewed on a regular basis, that distractions result in mislearning, and that the set of examples and rules that are presented must be unambiguous (so that the learner will not learn a misinterpretation). The misinterpretation is perhaps the most important single consideration, because there is a very reliable rule that if the presentation is ambiguous, some students will learn an unintended interpretation. A simple example would be a presentation that showed all examples of the concept "red" as being round balls and all examples of "not-red" as squares. Clearly, this demonstration cannot teach the naive learner what red really means because the learner has

* This project was conducted by Engelmann-Becker Corporation and coordinated by Don Steely.

the option of concluding that the word "red" refers to the color or that "red" refers to the circular shape (or that something is called "red" only if it is both circular and red in color). The problem with presentations that present possible misinterpretations is that while students may perform perfectly on the initial examples (red balls, for instance), it is not until later that we discover that hey don't understand red.

A more sophisticated illustration of misinterpretation would be provided by a poorly designed series of examples used to teach main idea. Let's say that for the first four examples, the main idea is expressed by the first sentence in the passage. The students perform marvelously on these examples. The next example, however, may be one that contains no sentence that expresses the main idea. The prediction, based on this poor set of examples, is that when some of the students reach this last example, they will identify the first sentence as the main-idea sentence and that it will require great effort to teach them the real concept of main idea. The point is that these students are not behaving in an unreasonable way. The series of examples the teacher presented strongly prompted them to attend to the "first sentence," just as certainly as the red balls would teach some children that red means round.

The results of the program analysis were, at best, frightening. Table 1 summarizes the averages of the five basal programs for the teaching of main-idea in grades 4, 5, and 6. Note that the number of lessons and examples refer to a three-year period.

The asterisked items provide some indication of the lack of precision exercised by these programs. Item 1 indicates that only 14% of the examples are taught. An example is considered "not taught" if the question of the type asked about the example had not been presented in the last 50 teaching days. (These basal programs, as you know, are not divided into daily lessons. To compute the lessons, we counted the total number of pages presented over the 3-year period, divided the total by 480 [160 lessons a year times 3 years].) The resulting number provided an arbitrary, average number of pages that should be covered during a "daily"

Table 1. Program Analysis Results Across Programs

		Means Across Program	Ideal
* 1.	Percentage of examples taught	14	100
2.	Percentage of questions ambiguous and not taught	88	0
3.	Percentage of answers to questions that were misleading and wrong	12	0
4.	Percentage of minimum discriminations not taught	5	0
5.	Percentage of variation in teacher presentation wording	14	0-15
6.	Percentage of variation in student workbook wording	44	0-15
7.	Percentage of variation in items, teacher presentation	10	0-50
8.	Percentage of variation in items, student workbook	18	0-50
9.	Percentage of questions relevant to concept, teacher presentation	62	100
10.	Percentage of questions relevant to concept, student workbook	75	100
* 11.	Percentage of probability of correct interpretation	27	100
12.	Percentage of response variation	13	0-50
13.	Percentage of visual distraction, student workbook	25	0
14.	Percentage of academic distraction, teacher presentation, student workbook	31	0
15.	Percentage of strength of teacher presentation responses	89	100
16.	Percentage of strength of student workbook responses	72	100
17.	Percentage of prompted, teacher presentation	24	0
* 18.	Percentage of prompted, student workbook	49	0
* 19.	Day since two examples were presented	62	?
20.	Total number of examples in program	66	?
* 21.	Number of student examples on same day as teacher material	9	?
22.	Percent of student examples on same day as teacher material	14	?
* 23.	Total number of lessons	22	50-80
* 24.	Percentage of examples for which correction is specified	0	100

* Large discrepancy between program and ideal.

lesson.

Item 11 shows that the probability of a correct interpretation (based on the set of examples presented by the program) is only 27%. In other words, there are approximately 4 possible interpretations that are perfectly consistent with the set of examples presented by the program. Item 18 shows the percentage of prompted examples (49 percent). These are items that give the student the answer. Item 19 indicates that a period of 62 days elapses before two ore more examples of main-idea are presented in the program. Over the three-year period, only 66 examples of main-idea are presented, only nine of these appear on the same day in the teacher and student material, and only 22 lessons deal with main idea at all. No specific correction procedures were specified in any of the five programs (item 24).

This analysis of main-idea suggests that if the teacher follows the average program and teaches precisely according to the program specifications, the programs are incapable of teaching the average student. The student will be bombarded by spurious prompts, will possibly be confused by distractors and variation in teacher wording, will be misled by the set of examples the teacher presents, will receive practice that is sparse and poorly designed, and will receive ambiguous and confusing instructions from the teacher.

The analysis of other skills paralleled that of main-idea. Fact-versus-opinion, for example is frequently taught so that it is perfectly misleading. Fact and opinion are taught as exclusive categories, which means that a person could not have an opinion that was a fact. (John said, "It's Friday today." It's a fact that John said that it's Friday today. If its is Friday, it's a fact that it is Friday. It is further a fact that in John's opinion, it is Friday. The material provided by these basals does not typically make these distinctions. Instead, it suggests that if something is an opinion, it is not a fact.)

How the Teachers Teach

The programs are basically incapable of teaching the average student, but possibly the teachers embellish these programs with good teaching that makes them work for the students. Certainly, we've all heard talk from teachers about how they don't follow the program and how they improved on it. (We received the same kind of information from the teacher interviews, where the teacher indicated that they deviated from the program specifications about 20% of the time.)

Probably the most interesting fact about the performance of the teachers in the study was that not one teacher deviate in any way from the specifications for the primary part of the lesson. Teachers sometimes didn't do the enrichment or additional activities provided by the teacher's guide, but followed the lessons precisely. Note that they were given no instructions about how to present other than, "Just present the lesson they way you normally would."

The tapes of the teaching were analyzed two ways—they were first analyzed without referring to the instructional program; next they were compared with the specifications provided by the program.

The teaching provided by the teachers (regardless of the program used) was not sound from a technical standpoint. The following is a brief profile of how the average teacher in the study taught:

1. The *maximum* rate of the teachers' presentation produced an average of 4.2 responses per minute. On student-reading tasks., the maximum rate was slightly higher—4.6 responses per minute.
2. The teacher presented 84% of the tasks to individuals and 16% to the group.
3. The teachers gave the answer to 34% of the tasks, either by responding with the students or by modelling the answer.
4. The teachers praised nearly half of the correct student responses (46%). Most praise was directed to individual students (95%). Only 2% was behavior-specific praise, rather than general praise.
5. The teaching presentations produced a student error rate of 27%. Only 37% of these errors were corrected. Of those mistakes for which a correction was provided, only on 10% was the student retested to determine whether the information provided by the correction was actually communicated to the student.

Table 2 compares the average teaching behaviors with ideal teaching.

As mentioned earlier, all teachers followed the specifications that were provided by the programs they used. If we compare their teaching with the teaching that would have resulted if the program were presented by some kind of recording device, we do notice some differences, however. These differences are caused by one problem—student mistakes. The teachers responded to these mistakes, and when they did, it typically increased the number of questions that were judged irrelevant or ambiguous. For examples, students read a main-idea passage that does not contain a topic sentence that expresses the main idea. The students had just finished reading three passages in which the main-idea was expressed as the first sentence in the passage. The passage they read now indi-

Table 4.
Mean Percent At or Above Different Criterion Percent

	Criterion Percent Correct		
Topic	90%	75%	50%
Main Idea	10%	33%	58%
Key Words	8%	32%	65%
Map Skills	30%	33%	56%
Inferences	15%	30%	62%
Context Clues	0%	0%	15%
Relevant Details	24%	82%	99%
Cause Effect	10%	30%	60%
Fact/opinion	0%	25%	70%
Means across all topics	12%	30%	55%

experience, but are quite incapable of responding to the problems with effective remedies.

Student Outcomes

After the taping, students were presented with test worksheets that tested the material that had been presented during the taped lesson. Table 4 summarizes the student performance on 8 topics. These outcomes dramatically confirm that the programs are incapable of teaching if presented as taught, and that the teacher's presentation was technically poor and presented a sequence of tasks that was actually inferior to that presented by the printed program. Although there was some variability from topic to topic, the tests disclosed that the students did not generally understand the concepts and information the teacher had just presented. The three topics that are of most interest to traditional educators are *main idea, context clues,* and *inferences.* No more than one-third of the students taught these topics scored more than 75% correct on what the teachers had just finished teaching. When we consider all the topics that were tested, we see a very frightening trend. Only about one-half of the student's scored 50% correct on the material just presented.

The first response to these results is perhaps shock. Imagine only about 30% of the students understand even 75% of what the teacher is trying to convey. When we look as the results in a broader context, however, we may draw the conclusion that the results are the inevitable outcome of traditional education Consider achievement tests. Items for these tests are designed so they will maximized individual differences and "spread the distribution." The test designers achieve this spread by designing items that are passed by about half the children (so the average child will correctly answer about half the items). The same pattern of correct responses appears in the results of the tests for the various topics. The average student correctly responds to about half the items. The basal programs, therefore, seem to be quite consistent with the achievement tests that are used to evaluate programs; however, neither the programs nor the traditional evaluations are appropriate for good instruction.

Table 5. Teacher Reports on Main Idea and Student Performance

T:	What percent of the students should master any skill?	86%
S:	Percent of students at 90% criterion on all topics	12%
	Percent of students at 75% criterion on all topics.	30%
T:	What percent of the students could do the workbook exercises after the lesson was taped?	72%
S:	Percent of students at 90% criterion level on all topics	12%
T:	What percent of the students need more practice on the topic taught?	58%
S:	Percent of students below 75% criterion level on all topics	70%
T:	How much practice do they need?	1 week
S:	Percent of students below 50% criterion level on all topics	55%
T:	What percent of the students master main idea?	56%
S:	Percent of students at 90% criterion on main idea.	10%
	Percent of students at 75% criterion on main idea.	33%
T:	What percent of students remain unchanged?	40%
S:	Percent of students below 75% criterion on main idea.	67%
T:	How deficient is the program for teaching students main idea?	16%
S:	Percent of students below 75% criterion on main idea.	67%

Teacher Verbal Responses

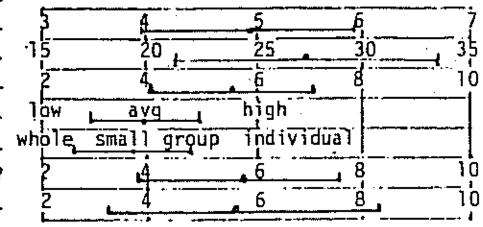
The reports by teachers generally showed that the teachers were not familiar with the details of the program they used, were not greatly aware of their teaching behavior, and greatly overestimated their students' understanding of the material presented. Table 5 gives a summary that compares their verbal responses to seven questions on student mastery on the topic *main idea*.

The final step that we took in this study was to determine the extent to which the teachers we sampled were typical of a broader population of teachers. To make this comparison, were designed a questionnaire that was sent to 3,000 teachers in grades 4, 5, and 6. The same questionnaire had been presented to the experimental teachers as apart of their first interview. Sixteen percent of those receiving the questionnaire responded (493 responses). The responses provided by the experimental teachers showed that the teacher gave atypical responses on 12 of the 94 scoreable items on the questionnaire. The experimental teachers, in other words, seemed to be a representative sample of teachers who were interested enough in instruction to re-

Table 6.

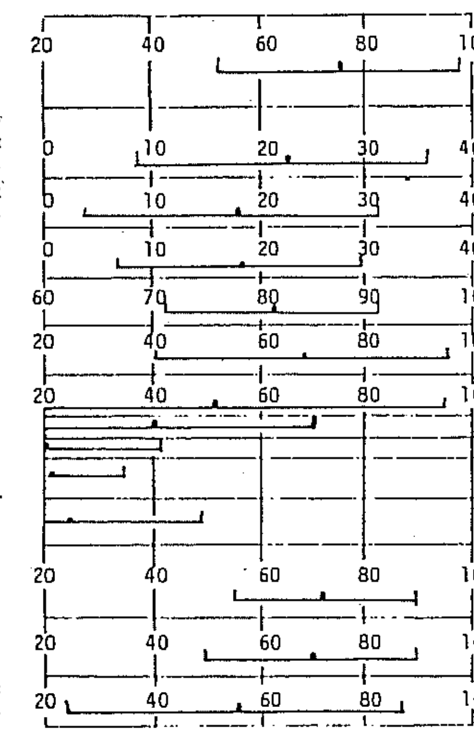
Section 1

1. What reading program do you use? (If you use more than one, list the one you use most and answer the questions based on that program.) ....
2. What grade do you teach? ....
3. How many students do you have for reading? ....
4. How many years teaching experience do you have? ....
5. How would you describe your reading class? ....
6. How would you describe the teaching instruction in your class? ....
7. How many hours per week do the students spend in reading instruction? ....
8. How many hours per week do you spend actually teaching reading? ....



Section II

1. In oral teacher directed activities how closely do you follow the procedures that are specified in the teacher's guide? ....
2. If a good student makes a mistake on an orally-presented reading activity, what percentage of the time will at least one other student repeat the same mistake? ....
3. If a low student makes a mistake, what percentage of the time will at least one other student repeat the same mistake? ....
4. During an average lesson, what percentage of the student responses will be incorrect? ....
5. As a generally acceptable guide, what percentage of the students *should* be able to master any particular skill in this reading program? ..
6. What percentage of the mistakes made during oral presentations do you correct? ....
7. Indicate the percentage of time you use each of these corrections:
   - Call on another student ....
   - Tell the student the answer is wrong and repeat the question ....
   - Tell the answer and then repeat the question ....
   - Tell the answer, repeat the question, then repeat the question to the whole group ....
   - Ignore the mistake and permit the students to discover a natural consequence of their error ....
8. Most reading programs make general comments about praising and reinforcing students for correct responses. What percentage of the student responses do you reinforce in some way? ....
9. What percentage of the time do you call on individual students to answer questions (rather than the whole group or class)? ....
10. In what percentage of the lessons do worksheet exercises for a topic appear the same day as teacher activities for that topic? ....



11. There are places in the program where it says to help or guide the students. In what situations is this appropriate? (Check all that apply) .

Students have just learned the skill	79%
Students rarely use the skill and tend to forget it	55%
Students need the help or they will never master the skill	59%
Students have never really been taught the skill	54%
Students must answer an open-ended question	35%

12. How would you describe the teacher directions of your program?

General statements of what to do	19%
Specific steps you are to follow	19%
Specific steps you are to follow and some direction on what to say	47%
Detailed scripts of exactly what to say and do	15%

turn the questionnaire. Table 6 gives part of the questionnaire, with the means for each item marked with a dot, and the range of responses indicated with a bar. If you want to compare yourself to the average traditional teacher in grades 4 through 6, answer the quests and compare your responses with the dots.

A Final Word

This study made me feel very sad, not so much because the results surprised me, but because the tapes of the teachers revealed both concern and a lot of raw talent. Most of the teachers who volunteered for this study were clearly intelligent people who were trying very hard to do an important job. Their verbal responses and the questionnaire responses suggest that these teachers are quite aware of the more obvious learning problems that their students experience. They know, for example, that students tend to confused the title or the first sentence with the main idea. They simply don't know how to avoid this problem, how to teach in a way that will help solve it; and how to provide explanations and examples that correct the problem. As it is, their talent, their potential to be super-teachers, is unfulfilled, in the same way that the potential of their students is. ◆

Table 6, continued.

13. What kind of mastery testing is specified in the program? ....

Repeated items from student workbook	5%
Items similar to those in student workbook	58%
Items not similar to those in student workbook	5%
Standardized tests specified	16%
No mastery testing specified	9%

14. What kind of follow-up activities are specified for students who do poorly on the mastery tests? ....

No follow-up exercises specified	23%
Lessons to be repeated	13%
Supplemental exercises, the same for any student below mastery level	34%
Supplemental exercises, different for different mastery test results	21%
Lessons to be repeated with specified criterion for those lessons	4%

Section III

1. About how many lessons in the level you teach focus on this skill? 0 10 20 30 40
2. About how many minutes long are lessons or parts of lessons that focus on the main idea? 0 10 20 30 40
3. What percentage of the students knew main idea before the lessons? .. 20 40 60 80 100
4. What percentage of the students master main idea? 20 40 60 80 100
5. Here are some typical types of main idea exercises. Put a P by those that are in the program you use. Put a U by those that you think are particularly useful in teaching main idea. You may mark both P and U for any items .... P U UP
   - Students underline main idea sentence when it is topic sentence.
   - Students decide if first or last sentence tells the main idea.
   - Students indicate which of three sentences tell the main idea.
   - Students decide if paragraph contains a main idea sentence and, if so, they underline it.
   - Students write their own main idea sentence.
   - Teacher goes through paragraph one sentence at a time and asks if that sentence tells the main idea.
   - Teacher asks students to tell the main idea and then has students try to find a main idea sentence in the paragraph.
6. Indicate the percent of students that have those problems:

• confuse main idea with title	30%
• think that main idea is first sentence	42%
• think main idea must be a sentence that comes from the text	46%
• cannot pick out the correct main idea sentence if it does not contain any phrase from the passage	42%
• cannot generate main idea sentence	36%
• other	24%

7. How closly do you follow the specified directions in your program for teaching main idea? 20 40 60 80 100
8. Indicate how you generally change your program's specific activities for teaching main idea. (Check all that apply.) 20 40 60 80 100
   - I skip the following percentage of specified activities: ....
   - I add the following percentage of activities: ....
   - I change the following percentage of specified activities: ....
9. If you change specified activities, how do you change them? (Check all that apply) ....

Provide more directed teaching	71%
Provide more review	55%
Do more testing	7%
Give more examples	79%
Give harder examples	25%
Give easier examples	32%
Spend more time on teaching the main idea	59%

Videodisc Instruction Teaching Fractions to Learning Handicapped and Remedial Students

by Bernadette Kelly
Douglas Carnine
Russell Gersten
Bonnie Grossen

From *ADI News* Volume 5, Number 3 (Spring, 1986)

The National Assessment of Educational Progress reported that, nationally, "performance of fraction computation is low, and students seem to have done their computation with little understanding" (Lindquist, Carpenter, Silver, & Matthews, 1983). For example, the assessment found that only one-third of the U. S. seventh-graders can add $\frac{1}{2}$ and $\frac{1}{3}$. The problem is even more pronounced for handicapped students.

Research on effective instructional practices with special education students give some clues about how to improve instruction. Englert (1984) measured mildly handicapped (M. H.) students' growth on a range of basic skills measures and correlated this growth with observed teacher performance. More effective teachers (classified on the basis of high student academic gain) *provided appropriate academic feedback to student errors more frequently* than did less effective teachers. The more effective teachers also *maintained pacing and higher student success rates* throughout each lesson. This set of variables has been found to be effective with low performing students in regular classroom settings (Good & Grouws, 1979; Gersten, Carnine & Williams, 1982; Rosenshine, 1983).

However, improved teacher training and improved teacher presentation techniques may not be enough. The curriculum itself is being called into question. A report from the National Council of teachers of Mathematics on National Assessment said, "It is necessary to reconsider the when, how, and what of the fractions sequence; some topics may need to be introduced earlier, others may need to be approached differently" (Carpenter, Coburn, Reys & Wilson, 1976).

This study compares the effectiveness of a traditional basal approach to teaching fractions with an innovative videodisc curriculum designed to teach basic fractions skills. *Mastering Fractions* (Systems Impact, 1985) incorporates sophisticated principles of curriculum design (Engelmann & Carnine, 1982) and harnesses the capabilities of the videodisc. The basal program *Mathematics Today* (Harcourt Brace Jovanovich) was selected from four widely adopted basal as the text most similar to the videodisc program in terms of five aspects of curriculum design: Lesson structure, review procedures, discrimination practice, example selection, and the use of clear, explicit strategies.

However, important instructional differences remain along those dimensions. The next section compares the two programs and cites research conducted on those dimensions. For a fuller discussion and additional research findings, see Gersten and Carnine (1984), Silver, Carnine and Stein (1981), and Stein, Jenkins and Carter (1983).

Comparison Between the Videodisc and the Basal Approaches

Lesson Structure

Each *basal* lesson beings with an introduction, followed by an explanation of the student-book pages, and then independent work. This structure results in students working independently for the last part of each lesson. Long periods of independent work may give rise to student inattentiveness (Gall, Gersten, Grace & Erickson, in press).

Mastering Fractions has a short explanation followed by problems that students work. This pattern of explanation followed by student written activities is presented for several concepts in each lesson. By presenting explanations with questions periodically within each lesson, more students remain attentive. Independent work is done in shorter, more frequent, segments to increase the amount of academic engaged time.

Review Procedures

In the basal program a skill is introduced and practiced, but then "disappears" for several days. For example, *Mathematics Today* teaches multiplication of fractions in one lesson. In subsequent lessons, other skills are introduced, including multiplication of whole numbers and fractions, and multiplication of mixed numbers. However, for the next three lessons students work with word problems, reciprocals, and division, after which students are expected to perform the multiplication of fractions independently on review and test lessons.

In *Mastering Fractions*, the skills of multiplying fractions is introduced and then practiced on *every subsequent lesson in the program*. Each new skill that is taught is reviewed cumulatively, or else incorporated into more complex skills.

Discrimination Practice

Students who learn to carry out certain steps again and again on the same type of problem may have difficulties when they encounter different problem types mixed together on a test. For example, a 14-day unit in the basal program introduces adding and subtracting fractions. In the next unit, students learn the strategies for multiplying and dividing fractions. No practice is given on *discriminating* between the strategies (e.g, multiplication and addition). In the review and test lessons, the programs types are still separated. Students never receive discrimination practice between strategies. After the two units, fractions operations do not appear again in the text *for the remainder of the school year*.

In *Mastering Fractions*, a skill is introduced, practiced, and within a few lessons mixed with other types of problems. For examples, exercises in the lesson presentation specifically address the differences between addition and multiplications strategies. If students have difficulty making the discrimination, specific remediation is given, after which students are required to work a set of problems involving both operations. The skills are then integrated with other types of problems on every worksheet.

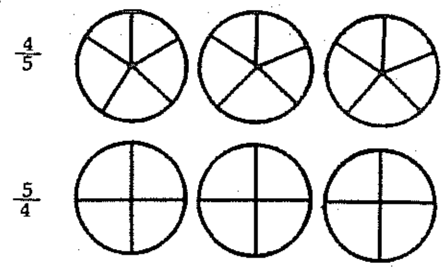
Darch, Carnine and Gersten (1984) compared the effectiveness of a regular basal mathematics curriculum with a curriculum program similar to *Mastering Fractions* in that it incorporated systematic discrimination practice. Students who received discrimination practice performed significantly better than students who did not on a criterion-referenced posttest and maintenance test. Englert (1984) also emphasizes the importance of discrimination practice for mildly handicapped students, to avoid confusion between related concepts.

Example Selection—Range of Examples

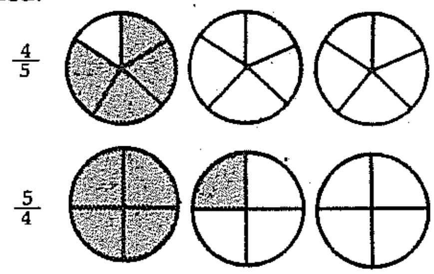
In the basal program, when students first encounter pictures of fractions, all examples are less than one. In the next grade level, mixed numbers are introduced as a whole number and a fraction, reinforcing the misconception that fractions can only represent qualities less than one. Improper fractions do not appear until the next grade level. A common error occurs when improper fractions are finally introduced; students represent these fractions as less than one; e.g., for $\frac{5}{4}$ students write:

Mastering Fractions teachers students a strategy for reading and writing both proper and improper fractions from the beginning of the program:

1. The denominator tells the number of parts in each group:



2. The numerator tells the number of parts used or shaded:



The wide range of examples prevents students from forming misconceptions and gives students a more complete understanding of what a fraction represents.

In a carefully controlled experiment, Carnine (1980) demonstrated how a limited range of examples can cause students to form misconception. The instructional task was to write fractions of a hundred as decimals. One group of students was presented with a wide range of examples, with numerators or one, two or three digits (e.g., $\frac{185}{100}, \frac{2}{100}, \frac{75}{100}$. The other group was presented with a limited range of examples; all numerators comprised two digits (e.g., $\frac{28}{100}, \frac{84}{100}, \frac{55}{100}$). Carnine hypothesized that students in the limited range group would learn the misconception that the decimal point is always placed directly in front of the digits in the numerator (i.e., $\frac{4}{100} = .4, \frac{185}{100} = .185$). His prediction that these students would not be able to generalize to other examples was verified. Students in the limited range groups scored 0% and 7% respectively on the problem types $\frac{X}{100}$ and $\frac{XXX}{100}$ on the immediate posttest. Students who had received the full range of examples scored 89% and 93% respectively.

Easily Confused Labels

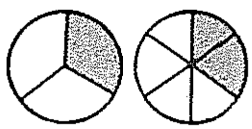
When highly similar terms (e.g., the terms numerator and denominator) are introduced at the same time, there is an increased likelihood that students will become confused. In the basal program, the terms *numerator* and *denominator* were introduced together in the same lesson. In subsequent fraction examples, the teacher referred to the terms *numerator* and *denomi-*

nator, and the labels appeared on some worksheets, but no systematic teaching ensured that students could successfully apply the labels to the appropriate parts of a fraction.

In the *Mastering Fractions* program, the introduction of the terms numerator and denominator were separated by several lessons, so that students were facile with one label before the other, similar label was introduced. This procedure decreases the likelihood that students will become confused and make reversals.

Explicit Strategy Teaching

In the basal program, students are not always given an explicit strategy to solve a problem. This could lead to student misunderstandings. Equivalent fractions serve as an example. In the first set of basal exercises, pictures of the two equivalent fractions, and three of the four fraction numbers are given; the students just count the number of shaded parts to complete the problem:



$$\frac{1}{3} = \frac{7}{6}$$

Students can write the fourth number and complete the equation without understanding anything about equivalent fractions. The students just count the shaded parts and write the numerator. In the final set of exercises given that day, the pictures are removed.

$$\frac{3}{4} = \frac{7}{8}$$

The student workbook say, "You may draw a picture to help you." At least some students will not be sure how many parts to draw or shade; unless, of course, they already know how to rewrite $\frac{3}{4}$ as $\frac{6}{8}$.

In *Mastering Fractions*, the strategy for equivalent fractions emphasizes this rule: When you multiply by one you don't change the value. When a fraction is multiplied by a fraction equal to one, the original fractions is equivalent to the new fraction: i.e.,

$$\frac{1}{2} \times 1 = \frac{1}{2}$$

$$\frac{1}{2} \times \frac{4}{4} = \frac{4}{8}$$

$$\text{so,} \quad \frac{1}{2} = \frac{4}{8}$$

With this conceptual basis for equivalent fractions, students are introduced to the strategy for figuring out the missing number, given a problem; e.g., $\frac{2}{3} = \frac{\square}{6}$. Students identify the fraction of one they must multiply $\frac{2}{3}$ by two to end up with 6ths. $\frac{2}{3}\left(\frac{\square}{2}\right) = \frac{\square}{6}$. The denominator of the fraction inside the parentheses is 2, so the fraction equal to one is $\frac{2}{2}$. Therefore: $\frac{2}{3}\left(\frac{2}{2}\right) = \frac{\square}{6}$. Thus, the missing numerator is 4. Therefore: $\frac{2}{3} = \frac{4}{6}$.

Kameenui, Carnine, Darch, and Stein (in press) compared a basal approach to introducing fractions with a strategy-based approach similar to that found in the *Mastering Fractions* curriculum. For the explicit rule-based strategy group, the teacher demonstrated concepts and skills in a step-by-step fashion. Teacher guidance was gradually and systematically faded until students were performing independently. Correction procedures directed students to the explicit instructions they had received. In contrast, the basal approach was much less structured. Emphasis was placed on activities using students discussion and the use of manipulatives. Students in the explicit strategy group performed significantly higher on a criterion-referenced posttest and on a transfer test of related fractions skills.

The Videodisc Technology

Videodisc technology has great potential as an instructional medium (Hofmeister, Engelmann, & Carnine, in press). One side of a videodisc contains 54,000 high resolution individual frames. The frames can be shown in rapid succession to create motion sequences or displayed as single frames for any period of time. Moreover, a teacher using a videodisc program has almost instant access to any portion of the disc. Using a remote control pad (very similar to the remote control for a TV) the teacher can access anywhere on the disc in a matter of seconds. Automatic stops can also be built into the disc; the program can then freeze on any predetermined frame allowing the students to work problems or the teacher to elaborate on a concept.

The *Mastering Fractions* program takes advantage of the videodisc medium to demonstrate concepts clearly. For example, when equivalent fractions are taught, a fraction is put on a balance beam. The side with a fraction tips down. When an equivalent fraction is placed on the other side, the balance becomes level. The video sequence shows what equality means in a vivid, compelling manner. Computer graphics, sound effects, highlights and other techniques also help maintain students attention.

The capabilities of the videodisc can do more than create compelling motion sequences. The videodisc can also assist the teacher in diagnosing and remedying student errors. Quizzes and tests on the disc help the teacher diagnose students having difficulty with a particular skill area. Following each quiz, addresses (numbers) are displayed on the screen for the skills tested. The teacher enters the address for the segment that is needed, providing immediate remediation, through demonstrations and extra practice problems.

Method

A study was conducted to determine whether the instructional features incorporated into the videodisc program would have a significant effect on student performance. The study compared the relative effects of *Mastering Fractions* and a traditional basal program on student acquisitions of skills in a unit on fractions. Classroom behaviors known to be correlates of learning (academic engagement and success rate during the lesson) were also measured, and an analysis of students' error patterns was made. Student attitudes were also assessed, and information on obtained levels of implementation were recorded.

Subjects

Prior to training, subjects from two high school math classes were screened for: (a) mastery of the preskills necessary to learn basic fraction concepts and operations, and (b) prior knowledge of the specific skills to be taught. One was a remedial math class containing 22 students, including 11 special education (mildly handicapped) ninth and tenth grades. The other general math class contained 12 ninth-graders in need of remedial math, along with 6 ninth, tenth, and eleventh grade mildly handicapped students. In each classroom, qualifying students were randomly assigned to the basal text (BT) or interactive videodisc (IV) treatment. This resulted in four instructional groups. In the remedial class, 9 students were assigned to each treatment. In the general math class, 8 students were assigned to each treatment. Out of 34 subjects, only 28 completed the study and took the posttest; 26 students took the maintenance test. Table I shows the number of subjects qualifying in each group who completed the study.

Table 1..
Distribution of Subjects in the Four Instructional Groups

	Remedial Math	General Math	Total
Basal			
MH*	5	5	10
non-MH	3	3	6
Videodisc			
MH	6	1	7
non-MH	1	4	5

* mildly handicapped

Measures

Preskills screening test. A screening test, developed by the experimenter, was administered to ensure that students had mastered the requisite whole number skills for a unit in fractions (i.e., facility with basic addition, subtraction, and multiplication facts). The first part of the test comprised ten of the more difficult facts. All students who were tested achieved at least 80% and were eligible for the study based on this criterion.

The second part of the screening test was criterion-referenced to the skills to be taught in the fractions unit. Students who scored about 50% on this part were ineligible for the study. Ten students were excluded based on this criterion. Eligible subjects were grouped in pairs, matched on Total Math scores from the California Achievement Test and on pretest scores. Individual students within each pair were then randomly assigned to the two treatment conditions. the mean scores on a 6 item pretest for the videodisc and basal groups were 2.4 and 2.1 respectively. (Information was not available for all students; $N = 10$ for the IV group, $N = 14$ for the BT group.)

Measures of achievement. The principal measure for the study was a criterion-referenced test developed by the experimenter. Two parallel forms were developed as a posttest and a two week maintenance test. The test included the following skills, taught in both the IV and BT conditions: writing fractions from pictures, vocabulary (e.g., denominator), addition and subtraction and fractions with like denominators, multiplication of fractions, and multiplication of a fraction and a whole number.

Field-test versions of the CRT were given to 30 fourth and fifth graders who had had some fractions instruction. Internal consistency reliability was assessed for each form; coefficient alpha reliability was .98 for the posttest and .98 for maintenance test. Alternate form reliability was also evaluated; the Pearson correlation coefficient was .96.

Measures of classroom variables. Two classroom variables associated with higher student achievement are: (1) total time students are *actively engagement* during instructional activities (*time on-task*; Rosenshine, 1983), and (2) student *success rate* while doing independent seatwork (Fisher et al., 1980).

An observational recording form was designed to measure *active engagement* during instruction. There were two groups of students in each condition. Each group of students was observed either 3 or 4 times during the study. Student behaviors were recorded with a six-second, momentary time-sampling procedure. On-task behaviors included answering questions, writing, and attending to the lesson presentation. Behaviors recorded as off-task included doodling, sleeping, or chatting to another student. Other behaviors (e.g., passing out papers, waiting for teacher assistance) were recorded as transitional activities.

Students' independent seatwork was collected at the end of 4 observation lessons to measure *success rate*. The percent of problems attempted and the percent that were successfully completed were calculated.

Student responses on the posttest and maintenance test were analyzed to determine patterns of frequently occurring or chronic errors. The error analysis was used to pinpoint aspects of the treatments that could have contributed to student errors.

Measures of implementation. Implementation checklists were used to identify those elements of the teaching models that were consistently implemented and those implemented at lower levels. The checklists were similar to the form developed by Good, Grouws & Ebmeier (1983). All items on the checklists were operationally defined. Below are two sample items that were applicable to both instructional models.

	Yes	No	NA
1. Did the teacher award points for independent work done on the previous day?			
2. Did the teacher circulate during independent work reinforcing appropriate behavior?			

Items relating specifically to the IV model (e.g., whether the teacher checked student performance at the specified points in the lesson, or administered a daily review quiz) were developed using the videodisc teacher's guide. Items applicable only to the basal text method (e.g., whether the teacher provided an opportunity to use manipulatives, or whether the teacher supplied examples in addition to those presented in the text) were developed using the basal text teacher presentation book. Each item scored in the "Yes" category by the observer was tallied, and the percent of total checks possible was calculated for each lesson observed.

Measures of student attitudes. A questionnaire was developed, based on the work of Fennema and Sherman (1976). Students were asked their opinion on a 3-point scale in response to a series of statements that related to the students' evaluation of their math ability and of the relevance of fractions for daily life. For instance,

1. I think I could handle more difficult fractions.
2. Learning fractions is a waste of time.

Items were read to students one at a time and the question asked, "Is this true for you?" Students responded to each item with: Yes, No, or Don't Know.

Materials

Interactive videodisc. The materials required for implementation of the IV fractions curriculum were: A videodisc player, the videodiscs, consumable student worksheets, and teacher answer keys.

Each videodisc lesson took approximately 30 minutes to complete. Lessons typically began with a brief quiz covering the essential skills introduced in the previous lesson. The lesson presentation followed next—an explanation followed by written problems for each of several skills. After completing the lesson, students were assigned independent problems for seatwork. The worksheets comprised 20 to 40 items, including a variety of skills that students had learned thus far.

In the IV curriculum, every fifth lesson was a test. Teachers used the tests to determine whether a review of particular skills was necessary from any of the four lessons preceding the test lesson.

Basal text. The materials required for implementation of the BT fractions curriculum were: A teacher presentation book (with answer keys), student textbooks, and consumable worksheets. In some lessons, manipulatives were also used, e.g., paper strips or fraction pie models.

Each 30-minute lesson was designed to teach a single objective. Each lesson began with an introduction, in which the teacher used discussion and demonstrations to develop ideas. Next, the teacher guided students through several examples in the student textbook before assigning in-class problems. After completing the lesson, follow-up activities, usually involving manipulatives, were used to consolidate the concept developed in the lesson. Students were then assigned independent problems for seatwork. The worksheets comprised 20-40 items focusing on the student objective introduced that day.

Review tests were provided at the end of the unit, sampling each of the major skills and concepts that had been introduced. Teachers used the results of the review test to reteach concepts and skills that students had not mastered. The unit test was presented the next day. The review and unit tests sampled the same skills in the same order, and had a standardized test format.

Procedures

The teachers were the experimenter and a research assistant from the University of Oregon. Each teacher taught one condition for one-half of the study, then changed conditions for the remainder of the study.

Monitoring implementation. The teachers were observed on 4 occasions to assess the level of implementation in each classroom. Teachers received specific feedback on their performance, using the Implementation Checklist (discussed under *Measures*). Throughout the study, teachers discussed any problems associated with the implementation of the two approaches.

Observers. Two trained observers recorded students' time on-task and percent correct responses on independent worksheets, on 3 or 4 occasions for each group of students. Before collecting the experimental data, the observers practiced using the instruments until interobserver reliability exceeded 85 percent.

Administration of measures. Criterion-referenced tests were administered to all students participating in the study immediately following the completion of the unit (posttest), and two weeks after completion of the unit (maintenance test).

Students' on-task behavior and success rate, and the levels of implementation were measured on the second, fourth, seventh, and ninth days of the intervention. The experimenter conducted student attitude surveys before and after completion of the study.

Results

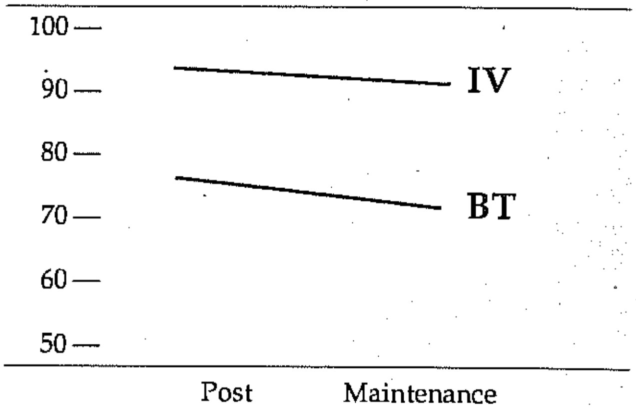
The primary dependent variable was student performance on the 12-item criterion-referenced tests (post and maintenance). A 2 x 2 analysis of variance (Anova) was performed on the CRT scores. The between-subjects factor was the instructional method (videodisc versus basal text); the within-subjects (repeated) factor was the time of test (post and maintenance). Significant main effects were found for the instructional method ($F = 17.28, p<.001$) and for time of test ($F = 4.53$, $p<.05$). There was no significant interaction. Thus, the effect was maintained over a 2-week period. Figure 1 shows the mean scores for students in each condition on the post and maintenance tests. The IV group scored at a clear mastery level and was about 20% above the BT group.

Students in the videodisc and basal conditions were on-task 96% and 84%, respectively, during observation periods. A Mann Whitney U –Test indicated a significant difference between the two conditions ($U = 3.5$, $p<.005$). Student's performance on independent seatwork was 92% for the BT group and 94% for the IV group.

Levels of implementation were extremely high in both conditions; 93% of the possible implementation behaviors were observed in the BT condition, and 94% in the IV condition.

Responses from the student questionnaires were summarized and assigned a score ranging from -1 (all negative responses) to +1 (all positive responses) for the students' perception of: (a) their ability to be successful in fractions, and (b) the relevance of fractions for daily life. Students in both conditions made

Figure 1.
Mean Percent Correct Scores on the Post Test and Maintenance Tests for Interactive Videodisc and Basal Treatments.



Standard Deviations

	Post	Maintenance
IV	.98	1.14
BT	2.19	1.88

similar gains in perceived ability; the IV group made larger gains in perceived relevance. The results are summarized in Table 2.

Table 2.
Mean Responses on Pre and Post Attitude Measures of Perceived Competence in, and Relevance of Fractions*

Competence	Pre	Post	Gain
IV	.08	.81	.73
BT	-.20	.58	.78
Relevance			
IV	.08	.57	.49
BT	.16	.31	.16

*Scores range from -1 (very negative) to +1 (very positive).

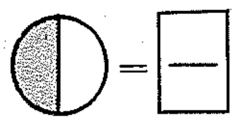
Discussion

The results of this experiment suggest that the different methods of instruction produce different levels of student mastery of the content covered. The students receiving videodisc instruction scored significantly higher both on the criterion-referenced posttest and on the maintenance test. The videodisc scores also dropped less dramatically over time—a non-significant drop of 1% compared to a drop of 7% for the basal text students.

While a significant difference was found between the two conditions for students' on-task behaviors, it should be noted that levels of on-task were high in both conditions. Students receiving the basal lessons were well motivated and actively involved during the lesson. This would imply that the quality of the IV curriculum—not merely the teaching procedures used in the study—was largely responsible for the differences in student performance.

Patterns of student errors also confirm the importance of the specific differences between the programs. For example, a large proportion (75%) of students in the basal treatment made errors when asked to write the fraction for a diagram representing a fraction greater than one. Given the diagram

56% of the basal students wrote $\frac{5}{6}$, even though all students could correctly identify



as $\frac{1}{2}$. The inability of 75% of the basal text students to *extrapolate* to fractions greater than one is a predictable consequence of all examples being less than or equal to one during the treatment intervention. In contrast, only 8% of the videodisc students, who had been exposed to fractions greater than one, exhibited this error on the post test. This parallels the results of the Carnine study (1980) cited earlier.

Advantages and Disadvantages of the Videodisc Medium

There are other important advantages resulting from using the videodisc medium in the classroom, apart from the instructional capabilities already discussed. First, the videodisc presentation frees the teacher from demonstrating at the front of the classroom, and enables the teacher to move among the students and monitor their performance.

Second, a well designed videodisc program can improve the quality of instruction provided by less confident (e.g., reassigned) teachers. Not only does the videodisc program provide clear initial demonstrations, but also provides frequent checks on student performance which can help the teacher diagnose student errors and select appropriate remediation procedures.

Third, the discs are highly durable. Surface scratches do not hinder the video or sound quality when the disc is played. The quality of the disc does not deteriorate over time. The durability of the disc and long lasting quality of the audio and video result from the laser technology. The laser beam reads the grooves that lie below a heavy coating of plastic.

The most obvious disadvantage of the videodisc medium—as with any new technology—is the cost. However, the cost of hardware has already dropped substantially. Also, presenting videodisc lessons to *groups of students* makes the technology more affordable. The combined cost of the hardware and software for a program such as *Mastering Fractions* is about the same as two Apple microcomputers and one or two inexpensive math software programs. If the videodisc is used 5 periods each day with classes of 20 students, 100 students are served each day. In contrast, 2 microcomputers used for 5 periods each day serve only 10 students.

The capability of the videodisc medium to incorporate state of the art instructional design features, together with its cost effectiveness, speaks to its potential as a powerful instructional tool. ◆

References

Carnine, D.W. (1980). Relationships between stimulus variation and the formation of misconceptions. *Journal of Educational Research*, 74(2), 106-110.

Carpenter, T.P., Coburn, T.G., Reys, R.E., & Wilson, J.W. (1976). Notes from national assessment: addition and multiplication with fractions. *The Arithmetic Teacher*, 23(6), 137-142.

Darch, C., Carnine, D., & Gersten, R. (1984). Explicit instruction in mathematics problem solving. *Journal of Educational Research*, 77(6), 351-359.

Engelmann, S., & Carnine, D. (1982). *Theory of Instruction: Principles and Applications*. New York: Irvington Publisher, Inc.

Englert, C.S. (1984). Effective direct instruction practices in special education settings. *RASE*, 5(2), 38-47.

Fennema, E., & Sherman, B. (1976). Fennema-Sherman mathematics aptitude scales. *KSAS Catalogue of Selected Documents in Psychology*, 6,(1), 31.

Fisher, C.W., Berliner, D.C., Filby, N.N., Marliave, R., Cahen, L.S., & Dishaw, M.M. (1980). Teaching behaviors, academic learning time and student achievement; an overview. In C. Denham & A. Lieberman (Eds.), *Time to Learn*. Washington, D.C.: USOE National Institute of Education.

Gall, M., Gersten, R., Grace, D., & Erickson, D. (in press). Determinants of effective algebra instruction: *Final Report*. Eugene, OR: University of Oregon.

Gersten, R. & Carnine, D. (1984). Direct instruction mathematics: a longitudinal evaluation of low income elementary school students. *The Elementary School Journal*, 84(4), 395-407.

Gersten, R.M., Carnine, D.W., & Williams, P.B. (1982). Measuring implementation of a structured educational model in an urban school district: an observational approach. *Educational Evaluation and Policy Analysis*, 4(1), 67-69.

Good, T.L., & Grouws, D.A. (1979). The Missouri mathematics effectiveness project: An experimental study of fourth-grade classrooms. *Journal of Educational Psychology*, 71, 355-362.

Good, T.L., Grouws, D.A., & Ebmeier, H. (1983). *Active Mathematics Teaching*. New York: Longman.

Harcourt Brace Jovanovich (1985). *Mathematics Today*. New York: Author.

Hofmeister, A.M., Engelmann, S., & Carnine, D. (in press). Observations from the development and field testing of an instructional videodisc program. *Journal of Special Education Technology*.

Kameenui, E., Carnine, D., Darch, C., & Stein, M. (in press). Two approaches to the development of mathematics instruction. *Elementary School Journal*.

Lindquist, M., Carpenter, T.P., Silver, E., & Matthews, W. (1983). The third national mathematics assessment: Results and implications for elementary and middle schools. *Arithmetic Teacher*, *31*(4), 14-19.

Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal*, *83*, 335-351.

Silbert, J., Carnine, D., & Stein, M. (1981). *Direct Instruction Mathematics*. Columbus: Merrill.

Stein, M., Jenkins, J.R., & Arter, J. (1983). Designing successful mathematics instruction for low performing students. *Focus on Learning Problems in Mathematics*, *5*(2), 79-99.

Systems Impact, Inc. (1985). *Mastering Fractions*. Washington, D.C.: Author.

Four Studies on Instructional Design Principles in CAI

by John Woodward
Doug Carnine
Russell Gersten
Mary Gleason
Gary Johnson
Maria Collins

From *ADI News* Volume 5, Number 2 (Winter, 1986)

Enthusiasm over computers and their potential impact on special education can be documented without much difficulty (e.g., Budoff, Thormann, & Gras, 1984; Blaschke, 1985). Such enthusiasm is only part of a general eagerness by many to see computers widely used throughout elementary and secondary education. While most advocates are adept at detailing the technical capabilities of this medium (e.g., immediate feedback, automatic scoring, individualized instruction), little has been done in the way of systematic research into the use of computers—in particular, computer assisted instruction (CAI)—for the mildly handicapped. This report summarizes four studies we have recently conducted in this area. They are the beginnings of what we consider to be systematic research into CAI for the mildly handicapped.

The limited research on the instructional effectiveness of CAI for handicapped and non-handicapped populations is complicated and often contradictory. After a comprehensive search of the literature, Forman (1982) concluded that achievement was rarely enhanced by CAI, even though students exhibited positive attitudes toward such instruction. We are not surprised by this finding, as little available software used in special education settings makes use of even the most rudimentary principles of sound instructional design and effective teaching (cf., Stevens & Rosenshine, 1981; Engelmann & Carnine, 1982; Brophy & Good, 1984).

In 1984, we began a series of CAI studies that examined different instructional design principles that have been articulated by Engelmann and Carnine (1982) and others. These principles have been empirically demonstrated as effective techniques in non-computer studies (e.g., Carnine, 1980; Carnine, Kameenui & Woolfson, 1982; Darch, Carnine & Gersten, 1984) and in our fifteen years of experience with Project Follow Through (cf., Stebbins et al., 1977). Through computer assisted instruction, we were able to isolate the effects of review cycles, size of teaching sets, explicit strategies, and correction procedures. We were able to do this in a variety of ways. Two of our studies compared popular commercial programs with software that we developed. In another, we examined the effect of one variable (a correction procedure) by modifying our version of the software. In the last study described in this report we used our software as an adjunct to a written curriculum to teach specific problem solving skills.

All of the studies described below were conducted with mildly handicapped secondary students. Students were screened for appropriate skill levels. For example, all students in the math word problems study were competent in basic operations through division and knew how to solve addition and subtraction word problems. Students whose skills were above or below this were not used in the study; those who remained were randomly assigned to conditions. Finally, in order to precisely measure academic development, tests were created for the particular skills taught in each study. The rationale and relevant details of each measure are described along with each study.

Vocabulary Instruction: Size of Teaching Sets and Cycles of Review

Many researchers, operating under the premise that word knowledge correlates highly with reading comprehension (Anderson & Freebody, 1981; Pearson & Gallagher, 1983; Tierney & Cunningham, 1984), have attempted to improve comprehension skills by teaching vocabulary. Unfortunately, those programs which were most successful in teaching new vocabulary also required the most time to accomplish that task. For example, a study by Beck, Perfetti, and McKeown (1982) attempted to teach only 104 words in 75 thirty-minute lessons. At the end of the study, students knew an average of 85 words that they did not know prior to the program, but this required 2,250 minutes of instruction or approximately 26 minutes per word. This amount of time is considerably more than that typically devoted exclusively to vocabulary instruction in the middle grades (Durkin, 1979; Roser & Juel, 1981). Computer assisted instruction, it would appear, offers the advantage of increasing instructional time on such a basic task without placing increased demands on a teacher's already limited time.

The purpose of this study was to compare two methods of computer assisted instruction (CAI) for teaching vocabulary to mildly handicapped adolescents. The study examined the effect of size of the daily teaching sets and provisions for daily and cumulative review on the acquisition and maintenance of word meaning. Two CAI vocabulary programs were used to present the same 50 words and definitions.

Two designs were used in this study: (1) a time to mastery (Will there be a significant difference between times required to meet mastery criterion on the 50 words by students taught with the two different CAI programs?) and (2) a fixed design, in which all subjects were tested after the seventh session. We also looked at differences between pretest and posttest scores as well as maintenance of effects two weeks after students achieved mastery.

Method

Twenty-four students were randomly assigned to one of the two CAI programs. Students worked individually on their assigned program 20 minutes a day for 11 days. All of the words, which were the same for both programs, were considered important by two or more special education teachers. A final list composed of 25 verbs and 25 adjectives was used.

The CAI programs. One program used in the study, the Small Teaching Set program, tests students on words and then creates lessons with the words they cannot identify (Carnine, Rankin, & Granzin, 1984). After testing the students on new words, the program provides instruction on a "teaching set" of no more than three words which the student missed on the test. Each lesson also includes a "practice set" with a maximum of seven words. The student must meet a specific mastery criterion on each word before it is removed from the practice set. The program tests the student on new words and adds words the student does not know to the practice set. Once the student has mastered ten words, the program presents a cumulative review lesson on those words.

The other program, the Large Teaching Set program, teaches words in sets of 25 words (Davidson & Eckert, 1983). The student may choose to see the words in any of four types of formats: (1) a teaching display which shows the word, its definition, and one example sentence; (2) a multiple choice quiz format; (3) an exercise in which a definition is displayed and the student must spell the correct missing word to complete a sentence; and (4) an arcade-type game in which the student matches words to their definitions.

Measures. A 50 item, multiple choice test was developed for the study coefficient alpha .79). This test was administered to all subjects as a pretest, as a criterion reference test at the end of seven sessions, immediately after mastery (or at the end of the eleventh session, and two weeks after mastery. There were also two transfer measures. One was a 10 item objective test in which students defined words orally. The other test required students to answer comprehension questions that require knowing the meaning of the words presented in several short passages.

Results

Eight of the 12 subjects (67%) in the Large Teaching Set program and 10 of the 12 subjects (83%) in the Small Teaching Set program met mastery criterion by the end of 11 sessions. The study was terminated after 11 sessions because the experimenter felt that the subjects who were still struggling to reach mastery were no longer benefitting from instruction. The mean number of sessions to mastery (for those who reached mastery) was 7.6 for those in the Small Teaching Set and 9.1 in the Large Teaching Set program. A t-test indicated this difference was significant (p less than .05). Hence, subjects in the Small Teaching Set program met mastery in significantly less time.

A 2 x 2 analysis of variance was performed on posttest and maintenance test scores, indicating no significant main effect for type of instruction. Results on the multiple choice test in the fixed-time design (i.e., the test administered to all students after seven sessions) indicates a slight, but nonsignificant difference in means favoring subjects in the Large Teaching Set program. Differences between scores on two transfer measures were also statistically nonsignificant.

Discussion

The unequivocal finding of the study was that the subjects taught with the Small Teaching Set program reached mastery criterion on the set of 50 words faster than subjects with the Large Teaching Set program. In addition, more students in the Small Teaching Set program reached mastery within 11 lessons. Given that the groups achieved equivalent levels of performance on the multiple-choice tests, their difference in acquisition rates becomes even more meaningful. Subjects taught with the Small Teaching Set program required less time to meet mastery criterion on the words, yet their posttest performance was equal to that of subjects in the other treatment who took longer reaching mastery. In addition, the shorter instructional time which the Smaller Teaching Set program subjects required did not negatively affect their retention of word meanings.

Math Word Problems: Explicit Strategies

Research into mathematical problem solving has been extensive. For example, Gorman (1968) identified 293 studies on word problems conducted between 1925 and 1965. In recent years, problem solving has been the subject of more research than any other topic in the mathematics curriculum (Lester, 1980). Yet the large number of studies has yielded little information for building effective interventions because of flaws in research design (Kilpatrick, 1978), vague descriptions of the experimental procedures (Silbert, Carnine, & Stein, 1981), and varying definitions of problem solving ability (Silver & Thompson, 1984). The success of further problem solving research depends less on a continued analysis of the learner and his or her deficiencies and more on: (1) an analysis of the limits of instruction the students are currently receiving, and (2) development of strategies that will work with low achieving students.

We investigated math word problems in order to determine whether handicapped students could learn to solve multiplication and division math story problems if taught a strategy that focused on how to choose the correct operation. It was hypothesized that students who received explicit instruction on choosing the operation would solve more problems correctly than students who received instruction that did not specifically focus on the choice of operation and concentrated, instead, on a more general strategy of manipulating units.

Method

A pretest-posttest design with random assignment of subjects to treatment groups was used to examine the effectiveness of two procedures for teaching mildly handicapped students to solve math word problems through computer assisted instruction. Twenty-six subjects were randomly assigned to two groups. The first group used a Direct Instruction math story problems program and the second used a highly regarded commercial program. Each subject worked at a computer for 15 to 30 minutes a day for 11 days.

The CAI Programs. The Direct Instruction program, Analyzing Word Problems, (Carnine, Hall & Hall, 1983) was based on principles of a theory of instruction articulated by Engelmann and Carnine (1982). The approach requires direct teaching of a clearly-specified step-by-step strategy. When a teacher is instructing the students, the teacher models each step in the process, heavily prompting the students as they continue to use the process, and then systematically fades the prompts until the students reach independence. When students make errors, the teacher again models or provides a prompt based on a previously taught rule. This style of strategy instruction was incorporated into the Analyzing Word Problems program. The program taught students how to solve multiplication and division word problems in a step-by-step fashion. When students erred, they were given a rule-based correction.

The Semantic Calculator (Sunburst Communications, 1983) was used as a contrast to the DI program. This program is based on the premise that the major difficulty in solving word problems comes from inappropriate manipulation of units (e.g., weeks, apples, dollars, etc.). If students could be taught to extract from the problem the quantities needed to solve the problem and the correct units for the answer, they would be able to solve the problem. In this program students were guided through story problems by answering "How many?" and "What?" questions about word problems that were written on worksheets. Next, the students used the letters A and B to type in the operation that should be performed on the numbers (i.e., "A/B" to divide and "A x B" to multiply) and predicted the units in the answer. The computer then told the student what units were used to express the answer to the problem. If the student answer did not match that of the computer's the student knew that he or she should go back and try again.

Measures. Both the pretest and posttest were 28-item tests comprised of 11 multiplication, 10 division, 2 addition, and 5 subtraction problems. all items were selected from three major arithmetic intermediate level textbooks and from the California Achievement Test. Sixty-eight percent of the problems on the test were like ones included in the instructional lessons; the remaining 32% were transfer problems.

Results

Results indicated no significant differences between performance of the Direct Instruction group and the Semantic Calculator group on the posttest and in the amount of time used to take the posttest. Interviews with students as they performed problems (i.e., choosing the correct operation and telling a reason for the choice) did yield a statistically significant difference between the groups favoring the Direct Instruction group, but the mean performance levels for both were not educationally significant.

Discussion

There are many possible reasons why there were no significant differences between groups. Eleven days at 25 minutes a day may have been too short of an intervention. With a longer treatment period, it would have been more certain that an unacceptable level of performance was attributable to other factors. Further, observations of student performance during the study indicated that many students typically ignored prompts on the screen that told them what to do next. Hence, through a failure to attend the students were missing opportunities to learn from their errors.

It is also conceivable that mildly handicapped students may need more teacher-directed instruction before using a computer for additional practice opportunities. The presentation of the problem solving strategy on the computer lacked the subtlety and flexibility that a teacher adds to instruction. Good teachers gather a considerable amount of information about how students are learning a new skill—particularly one as difficult as problem solving—and modify their teaching accordingly. Therefore, the most appropriate use of a computer for students such as these may be for guided practice (i.e., as a medium for reviewing material that is already familiar to the students).

Reasoning Skills: Correction Procedures

Much of the recent literature on improving special education teaching practices has stressed the importance of providing academic feedback to students when they make errors (Carnine, 1980); Reith, Polsgrove & Semmel, 1981; Stevens & Rosenshine, 1981). Furthermore, a recent meta-analysis of the limited research on corrective feedback by Lysakowski and Walberg (1981) suggests that detailed corrective feedback is superior to merely telling students whether their answers are right or wrong. Just telling students they are wrong (called a "basic correction") does not help them solve the problem correctly. These authors suggest that students need to see an overt model of all the steps necessary for an appropriate response. By observing a model of all the steps necessary in obtaining a correct response, students receive detailed information on how to solve the problem. This procedural knowledge should be of use when they encounter similar types of problems. This type of correction will be refereed to as an "elaborated correction."

The primary intent of this study was to examine whether remedial and mildly handicapped students who receive elaborated correction procedures would perform significantly better than students provided with basic corrections. We also examined the differences regarding acquisition time between students.

Method

Thirty-four students were randomly assigned to the Basic Correction or Elaborated Correction group. The Elaborated Correction group used an unaltered copy of the CAI program used in the study. The Basic Correction group used a modified version of the program. If a student in this group made an error, they were only given the correct answer. This was the only difference between the two conditions. In both conditions, students worked individually on a microcomputer. Students worked on their respective version of the program until they completed the lessons.

The CAI program. The Reasoning Skills program (Engelmann, Carnine & Collins, 1983) was designed to teach students two major objectives: (1) to draw conclusions from two statements of evidence, and (2) to determine whether a three-statement argument was logical or illogical. The program taught students about overlapping classes and non-overlapping classes. They learned that there are three possible key words (some, all, no); the same rule holds for all three. It also taught students relevant rules for constructing and analyzing arguments. The other major objective of the program was to teach students to identify unsound arguments. For logically unsound arguments, students were taught to specify one of three reasons why an argument was unsound.

Measures. The Test of Formal Logic (Collins, 1984) was the primary dependent measure in the study. The purpose of this test was to measure a student's ability to construct and analyze syllogistic arguments. Two alternative forms of the test were designed; Form A was used as the pretest and maintenance measure (given two weeks after treatment terminated) and Form B was used as the posttest measure (given immediately after the treatment). The internal consistency reliability (coefficient alpha) for Form A was .90. Parallel form reliability between Forms A and B was .84.

There was also a 15 item transfer test that evaluated subjects' abilities to generalize what they had learned on the computer to similar analytic tasks, but in prose paragraph form. The transfer test was devoted to the most difficult objective on the program—deciding whether arguments were sound, and, if not sound, giving a reason. The test was given to subjects on the day after they completed training on the CAI program.

Results

A 2 x 3 analysis of variance (ANOVA) with one between subjects factor (Type of Correction) and one within subjects factor (Time of Test) was performed on the data. This analysis involved a planned comparison that looked at the post and maintenance tests only. The ANOVA indicated a significant difference favoring the Elaborated Corrections group (p less than .001). There was also a significant difference between the two groups on the transfer test, again favoring the Elaborated Correction group (p less than .05).

Data were collected on the time students took to complete each of the five lessons. The purpose of this analysis was to see whether students in the Elaborated Corrections group took more time to complete the lessons. A 2 x 5 analysis of variance (ANOVA) with repeated measures was performed on the time-per-lesson data and a nonsignificant difference between groups was found.

Discussion

This study was the first to explore experimentally the effectiveness of elaborated corrective feedback in teaching a complex cognitive skill to handicapped learners. The results indicate this is an effective instructional procedure.

The roughly equivalent time for the two groups to complete the five lessons seems anomalous at first. With more text to read in elaborated corrections, that treatment would seemingly take longer to complete the lessons. Completion times were not significantly greater for the elaborated corrections group, however. The extra time to read the elaborated corrections may have been compensated for by faster acquisition of the material. In both versions of the program, the computer would return a student to items that were missed earlier in a lesson. If elaborated corrections resulted in fewer mistakes, students would spend less time returning to missed items. This interpretation suggests that taking more time early in a complex instructional sequence to offer elaborated corrections may, in fact, lead to savings in instructional time later in the program.

Both the basic and elaborated correction groups improved their reasoning skills as measured by the dependent variable. The groups demonstrated a mean score of 68-70% on the posttest (a dramatic gain from the mean scores of 26 to 34 percent on the pretest). The systematic design of instruction - particularly through a series of carefully controlled rules - may have contributed to this gain. Reasoning skills were acquired without any instruction from the teacher. Typically, CAI programs merely provide drill and practice exercises to supplement teacher instruction. Here the program was a true tutorial and did all the instructing.

Health Ways: Problem Solving Skills

Secondary students spend a considerable amount of their time completing application-oriented activities. These academic tasks often involve higher-order cognitive skills, and students are asked to make a variety of inferences about a subject area by prudently using facts, concepts, and strategies or problem solving skills. Some writers (Doob, 1972; Greenblat & Duke, 1975; Budoff, Thormann & Gras, 1984) have suggested that one way to enhance the higher-order skills of students is through educational simulations.

While much of the research has concluded that simulations are no more effective than conventional instruction, many of these studies have been plagued by fundamental weaknesses in research design (Fletcher, 1971; Pierfy, 1977). In the study below, we addressed many of these problems and created an instrument that reflected the problem solving skills actually taught in the simulation. Finally, we have addressed a curious feature of previous simulation research: the general reluctance to combine simulation instruction with conventional instruction.

In virtually all previous studies, simulations have been contrasted with conventional teaching methods. Only on a few occasions have simulations and conventional instruction been compared to conventional instruction alone. Nor is it clear in most of these studies what constitutes conventional instruction. One of our interests in studying simulations was to investigate how effective instructional practices could be used to enhance—rather than replace—secondary level instruction, not only in terms of their effect on basic fact and concept retention, but as they related to higher-order skills.

Method

Thirty students were randomly assigned to either the conventional or simulation condition. All students were instructed for 40 minutes per day for 12 days. The lesson consisted of two parts. The first part, called *structured teaching*, was identical for subjects in both student conditions. Instruction was conducted in a large group of 12 to 15 students for this part of each lesson.

At the end of the initial instruction, students separated into two groups—one which worked on application activities (the conventional group) and the other with the computer simulation (the simulation group). The conventional group worked in the resource room under the supervision of the resource room teacher, who presented these students with a variety of application or review activities.

Simulation students, on the other hand, were taught in a computer lab, each student working individually at a microcomputer. The 12-day course of instruction for these students was broken into three phases: initial modeling (3 days), guided practice on three simulation games (2 days), and independent practice with individual feedback from the instructor (7 days). *The CAI program. Health Ways* is a commercial software program developed by Carnine, Lang, and Wong for the Apple II and IBM PC jr computers. The *Health Ways Supplementary Curriculum*, developed by Woodward and Gurney, extended information presented in *Health Ways* and the original *Health Ways* teachers' guide. Material was taken directly from two widely used junior high school health textbooks. All of the information was rewritten to control for vocabulary and amount of new information. Clippings from newspapers, news magazines, journal articles, and health pamphlets were also used in this supplementary curriculum. The reading level of the curriculum is approximately sixth grade.

Measures. Students were assessed one day, two days, and two weeks following the instruction. On the first day, students' acquisition of basic facts and concepts about health taught in the curriculum was measured by the Health Ways Nutrition and Disease Test.

The first 20 questions of this test were solely from the written curriculum. The remaining 10 were questions over material that appears in both the written curriculum and the *Health Ways* simulation. Internal consistency reliability (coefficient alpha) of this measure is .84. On the second day, the students were given the Health Ways Diagnosis Test, an individually administered test measuring prioritizing skills. This test was a set of three written profiles and measured health related problem solving skills (i.e., the student's ability to detect important health problems facing an individual, identify and change related health habits, and control stress as it increased due to the health changes). The Health Ways Diagnosis Test has a test-retest reliability of .81. Two weeks after the instruction the students were again given the Health Ways Nutrition and Disease Test. This served as a retention measure.

Results

The data were analyzed in two parts. Major analyses involved a comparison of scores on the two measures for the handicapped students. Because the Health Ways Nutrition and Disease Test was used in both the posttest and maintenance (retention) phases, a 2 x 2 analysis of variance (ANOVA) was used. Secondary analyses compared the performance of the handicapped students in the experiment with a random selection on non-handicapped peers who were in health classes.

Subscales analyses. The 30 item test was broken into two subscales: (1) items reinforced by *Health Ways* simulation, and (2) items taught in the curriculum and not reinforced by the simulation. Separate 2 x 2 ANOVAs with repeated measures were performed on each subscale. The effect on items reinforced by *Health Ways* was significant (p less than .01). For those items not reinforced the comparison was not significant (p less than .06). This indicates that the simulation was an effective vehicle for reviewing material that had already been taught in the written curriculum.

T-tests performed on the Diagnosis Test demonstrated a significant difference between the two groups (p less than .001) in problem solving skills. Simulation students were better able to diagnose health problems, prioritize them as to their effects on an individual's longevity, and prescribe appropriate remedies.

In a supplemental analysis, a one-way analysis of variance (ANOVA) was used to compare the test performance of the conventional and simulation groups with students from regular health classes who did not participate in the study. The purpose of this quasi-experimental comparison was to extend the posttest analysis to students of a comparable age group who were also receiving health instruction. Again, scores from each section of the Health Ways Nutrition and Disease Test and the Health Ways Diagnosis Test were analyzed.

Total score on the Diagnosis Test showed significant differences between the groups (p less than .001). A Tukey post-hoc comparison indicated significant differences favoring the handicapped simulation students over those in the regular classroom (p less than .01) as well as a significant difference favoring the regular classroom students over the mildly handicapped students in the conventional group (p less than .01).

A significant difference also appeared between the groups on the reinforced subscale of the Nutrition and Disease Test (p less than .01). Tukey post-hoc comparison showed a significant difference between the mildly handicapped simulation group and the two other groups (p less than .05), favoring the handicapped students taught by *Health Ways*, but no difference on items not reinforced.

Discussion

The results of this study support the use of computer simulations in teaching material not easily taught by traditional means. Further, a structured approach in simulations, one where outcomes are specified and controlled, does produce significant educational results.

We infer from the results that the explicit strategy instruction used to teach the simulation students about *Health Ways* was a successful bridge to "less" direct instruction activities. Further, the superior performance by those in the simulation group over non-handicapped students from regular health classes shows that explicit strategies instruction is successful in teaching unstructured academic tasks that involve higher level knowledge structures or strategies. The two non-handicapped students who had the highest scores on the Diagnosis Test articulated a prioritizing strategy comparable to that given by several of the handicapped students. Thus, many of the handicapped students in the simulation group showed a conscious awareness of the strategies that they were using, as did the two non-handicapped students, who may have achieved their awareness in a more intuitive manner.

Conclusion

The results of these four studies indicate that properly designed CAI can be effective as an instructional medium. These findings are consistent with our non-computer research that we have conducted over the last 15 years. Using sophisticated, empirically-based design principles can make a considerable difference in the effectiveness of any instructional program. Yet another outcome of these studies is that they begin to identify—with much greater clarity—the role of the teacher and his/her instruction independently from the computer. This perspective deviates from original questions about computer assisted instruction (e.g., Is CAI more effective or efficient than conventional instruction?). It forces us to look closely at the intersection of the teacher, the academic task, and the stage of instruction.

The Vocabulary Instruction study, for example, demonstrates that a skill requiring considerable practice can be adequately taught on a computer. Such a task is time consuming for a teacher and can be handled effectively by a computer. Furthermore, there is very little variation as it moves from one stage of instruction to the next (i.e., from introduction and teacher modeling, to guided practice, and independent practice). Note, however, that the task, as it was defined in the study, was one of memorizing vocabulary words. We did not teach nor assume that students would necessarily learn how to use the words expressively or detect their meaning from context. This would have required a different analysis.

The Reasoning Skills program, a teacher independent tutorial, was successful at teaching a more complicated academic task: logical inferences. It might be argued that the particular academic skills taught in this program were more discrete than, say, math word problems or the subtle problem solving skills addressed in the Health Ways study. If this observation is correct, then computer assisted instruction—carefully developed with instructional design principles and field tested—can be an effective, "stand alone" form of instruction.

Math Word Problems, on the other hand, do not share the same task complexity as the syllogisms. In this study, we found that the best way to teach skills such as math word problems, notoriously difficult for mildly handicapped students, may be through teacher directed instruction first, with the computer used for guided practice. This conclusion is tentative, and we will conduct another study after we revise our strategy.

Finally, the *Health Ways* study strongly suggests that facts and concepts, which were preskills to the problem solving activities, can be efficiently taught in group instruction without computers. The computer can be an effective tool after the preskills have been introduced and explicit strategies for using the simulation have been taught by the teacher. In this sense, a complex task like problem solving can be effectively taught in the guided and independent practice phases of instruction. Thus, the best use of computer assisted instruction requires a careful look at the academic task, the stage of instruction, and the appropriate role of the teacher. ◆

References

Anderson, R. C. & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association.

Beck, I. L., Perfetti, C. A. & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*, 506-521.

Blaschke, C. (1985). Technology trends in special education. *T.H.E. Journal 12(6)*, 73-77.

Brophy, J., & Good, T. (1984). Teacher behaviour and student achievement. In M. Wittrock (Ed.) *Third handbook or research on teaching* (in press).

Budoff, M., Thormann, J., & Gras, A. (1984). *Microcomputers in special education*. Cambridge, MA: Brookline Books.

Carnine, D. (1980). Phonics versus whole-word correction procedures following phonic instruction. *Education and Treatment of Children, 3*(4), 323-330.

Carnine, D., Hall, T., & Hall, R. (1983). *Analyzing word problems*. Unpublished computer program, Eugene, Oregon.

Carnine, D., Kameenui, E., & Woolfson, N. (1982). Training of textual dimensions related to text-based inferences. *Journal of Reading Behavior, 14*(3), 335-340.

Carnine, D., Lang, D., & Wong, L. (1984). *Health ways*. Unpublished computer program, Eugene, Oregon.

Carnine, D., Rankin, G., & Granzin, A. (1984). *Learning vocabulary*. Unpublished computer program, Eugene, Oregon.

Collins, M. (1984). Test of formal logic. Unpublished doctoral dissertation, University of Oregon.

Darch, D., Carnine, D., & Gersten, R. (1984). Explicit instruction in mathematics problem solving. *Journal of Educational Research, 77*, 351-359.

Davidson, J. G., & Eckert, R. K. (1983). Word attack! (computer program). Ranchos Palos Verdes, CA: Davidson & Associates.

Doob, P. (1972). Prospects for simulation gaming in health planning and consumer health education. eric ed 070 596.

Durkin, D. (1979). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly, 14*, 481-533.

Engelmann, S., & Carnine, D. (1982). *Theory of instruction*, New York: Irvington.

Engelmann, S., Carnine, D., & Collins, M. (1983). DIAL reasing skills program (computer program.) Eugene, Oregon: Engelmann-Becker Corp.

Fletcher, J. (1971). The effectiveness of simulation games as learning environments. *Simulation & Games, 2*, 425-554.

Forman, D. (1982). Review of the literature. *The Computing Teacher, 9*(5), 37-50.

Greenblat, C., & Duke, R. (1975). *Gaming simulation: Rationale, design, and application*. New York: Halsted.

Gorman, C. (1968). A critical analysis or research on written problems in elementary school mathematics. *Dissertation Abstracts, 28*, 4818A-4819A.

Kilpatrick, J. (1978). Research on problem solving in mathematics. *School Science and Mathematics, 78*, 189-192.

Lester, Jr., F. K. (1980). Problem solving: Is it a problem? In M. M. Lindquist (Ed.), *Selected issues in mathematics education* (pp. 29-45). Chicago: The National Society for the Study of Education.

Lysakowki, R., & Walberg, H. (1982). Instructional effects of causes, participation and corrective feedback: A quantitative synthesis. *American Educational Research Journal, 19*(4), 559-578.

Pearson, P. D. & Gallagher, M. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology, 8*, 317-344.

Pierfy, D. (1977). Comparative simulation game research: Stumbling blocks and steeping stones. *Simulation & Games, 8*, 255-268.

Reith, H., Polsgrove, L., & Semmel, M. (1981). Instructional variables that make a difference: Attention to task and beyond. *Exceptional Education Quarterly, 2*(3), 61-71.

Roser, N., & Juel, C. (1981, December). *Effects of vocabulary instruction on reading comprehension*. Paper presented at the annual meeting of the National Reading Conference, Dallas.

Semantic Clculator (computer software) (1983). Pleasantville, NY: sunburst Communications.

Silbert, J., Carnine, D., & Stein, M. (1981). *Direct instruction mathematics*. Columbus, OH: Merrill.

Silver, E., & Thompson, A. (1984). Research perspectives on problem solving in elementary school mathematics. *The Elementary School Journal, 84*, 529-545.

Stebbins, L., St. Pierre, R., Proper, E., Anderston, R., & Cerva, T. (1977). *Educaiton as experimentaiton: A planned variation model*, Vol. IV A-D. Cambridge, MA: Abt Associates, 1977.

Stevens, R., & Rosenshine, B. (1981). Advances in research on teaching. *Exceptional Education Quarterly, 2*, 1-9.

Tierney, R. J., & Cunningham., J. W. (1984). Research on teaching reading comprehension. In P.D. Pearson (Ed.), *Handbook of reading research* (pp. 609-656). New York: Longman.

Why Johnny Still Can't Read

Flesch, Rudolph, *Why Johnny Still Can't Read.* New York: Harper & Row, 1981

From *ADI News* Volume 1, Number 1 (Fall, 1981)

Why Still Can't Read is a scathing indictment of what Flesch calls the "Look-and-say approach to teaching reading." The book begins by documenting the massive extent of illiteracy in this country. Flesch summarizes a U.S. Office of Education sponsored study which concluded that 23 million people between the ages of 18 and 65 could not read a want ad or a job application. Flesch attributes this huge illiteracy rate to the fact that 85% of the schools in this country use programs based on the look-and-say approach. Flesch does not equivocate in his conclusion that because schools are using the look-and-say approach instead of phonics, "America is rapidly sinking into a morass of ignorance [p. 1]."

Flesch does not deal in vague generalities. He specifically names the programs likely—and those *unlikely*—to be effective. The effective programs, referred to as "The Phonic Five" are Addison Wesley, DISTAR, Economy, Lippincott, and Open Court. All other major reading programs are included in a list entitled "The Dismal Dozen." Flesch considers all twelve entries on this list to be look-and-say programs.

In a chapter entitled "Look and Say Exposed," Flesch summarizes the research results on the two approaches. Since 1911, there have been 124 studies comparing the effectiveness of phonics-first and look-and-say programs. Every one of these studies showed results in favor of phonics-first. Flesch specifically discusses Project Follow Through and the favorable results the DISTAR programs produced.

These overwhelming results lead to the question, "Since phonics are clearly superior, why are 85% of the schools still using look-and-say?" In a chapter entitled "The Great Coverup," Flesch suggests that leading educators and publishers of look-and-say programs have conducted a thorough and on-going coverup of the research findings. Whenever a new exposure of the look-and-say approach is threatened, someone—usually a professor on the payroll of a look-and-say publisher—is sent into the breach to defend the system. This has been going on year after year for over 50 years until today even some of the most experienced people in the field are confused or uninformed about certain areas of their own profession [p. 40].

The remainder of the book is devoted to "the ten alibis" used by look-and-say educators to defend their approach. The alibis include such topics as: the child isn't ready to read, teaching phonics is merely teaching word calling, no one method is best, English isn't phonetic, and the child can't learn because of his home environment. Every Direct Instruction teacher has heard these arguments for the look-and-say approach. Flesch examines each of these excuses and offers very compelling arguments to refute each one.

The only major criticism of the book, from a direct instruction standpoint, is that Flesch seems to overgeneralize that any phonics-first program will be effective with all children. He states that if a school is using any of "The Phonic Five," children will not have problems learning to read. Flesch never looks at critical instructional variables such as rate of introduction of sounds, teaching of essential preskills, amount of controlled practice, and cumulative review.

Why Johnny Still Can't Read is an excellent resource for comparing phonics-first and look-and-say programs. Flesch points out very clearly and forcefully that any phonics-first program is much more likely to be effective than a look-and-say method. Hopefully, Flesch will begin now to look in more detail at a comparison of the phonic programs and their relative effectiveness with all children, even the lowest performers. ❖

Reviewed by Randy Sprick

Becoming a Nation of Readers
The Report of the Commission on Reading

Prepared by Richard C. Anderson, Elfrieda H. Hiebert, Judith A. Scott, Ian A.G. Wilkinson

Summarized by Wes Becker

From *ADI News* Volume 4, Number 4 (Summer, 1985)

This report was produced under the auspices of the National Academy of Education's Commission on Education and Public Policy, and sponsored by the National Institute of Education. The Commission on Reading was chaired by Richard C. Anderson, Head of the Center for the Study of Reading at the University of Illinois.

The report summarizes the research-based knowledge on reading instruction, identifies problems with current practices, and recommends possible solutions. The report recognizes that quality instruction involves many elements and that improvement in reading instruction will require changes in many elements of reading instruction, not just a few magical tricks. It is the first nationally-based report to actively support Direct Instruction practices since the ABT Associate Reports on the Follow Through outcomes.

The Skills Involved in Reading

"Reading is the process of constructing meaning from written texts. It is a complex skill requiring the coordination of a number of interrelated sources of information" (p. 7). An analogy is drawn between the process of reading and the performance of a symphony. While reading can be analyzed into subskills that can be built up one at a time, "real" reading takes place only when all the pieces are integrated into a smooth performance.

In reading, being able to say the words (decoding) "gives access to their meaning" (p. 8). Understanding what is read requires a very substantial knowledge base to *construct* viable interpretations of text. Because people differ in their knowledge bases, different interpretations of text often occur. Fluent decoding skills are essential to good comprehension. Second graders with the best comprehension scores are the ones who decode fast and accurately. When decoding skills are weak, the time spent in trying to decode interferes with the interpretation processes. Good readers are better than poor readers in pronouncing nonsense words as well (e.g., *cade*, *mot*, etc.). Evidence like this supports the importance of a fluid decoding base. The evidence shows that the average third grader can read aloud about 100 words per minute, while the rate of the poor reader is about 50 to 70 words.

Skilled readers have learned strategies for reading different kinds of material. Complex, unfamiliar material must be read differently than familiar material. Strategies should also vary with the purpose of the reading—say, for a test or for fun. Skilled readers have strategies for monitoring the process of reading and detecting problems to be solved—such as inconsistencies, unknown words, etc.

Learning to read is a lot of hard work and the practice required can become monotonous if the teacher does not work to maintain motivation. "Teachers whose classes are motivated are described as business-like but supportive and friendly" (p. 15). Teachers with motivated students "conduct fast-paced and varied lessons. Tasks are introduced with enthusiasm and with explanations of why doing them will help one become a better reader" (p. 15). Failure is not fun, and poor readers show the apathy that goes with failure. Good teachers have ways to motivate poor readers by praising steps of the process (describing what the student does right). Reading is a skill that improves with practice. Finding ways to increase practice can improve reading.

Early Reading

The Home

Reading begins with the development of oral language skills at home. The knowledge base developed before school is very important to progress in reading comprehension. Once in school, the home experience continues to be important in knowledge development. Vocabulary development in particular is seen to have an important base in how parents talk about experiences their children are having. The kinds of questions a parent asks can aid or impede the development of reasoning skills. Reading aloud to preschool children is seen as an important aid to children learning to read. Records or tapes with "follow along books" can also be helpful. Having access to pencil and paper or chalkboards, with encouragement to write the letters of the alphabet, etc., is beneficial. Parents can also be tutors and directly or indirectly teach their children to read. Parents can and do play a critical role in their children's development of reading skills. We need more active plans to encourage effective parental practices.

Kindergarten Reading Instruction

Those involved with Direct Instruction since the days of the Bereiter-Engelmann Preschool (1964) and through 17 years of Follow Through know well that kindergarteners can be taught to read and to enjoy their skills. Twenty-one years later, the *Commission on Reading* is (as far as we know) the first National professional group to endorse the teaching of a systematic approach to reading in kindergarten. "Based on the best evidence available at the present time, the Commission favors a balanced kindergarten program in reading and language that includes both formal and informal approaches. The important point is that instruction should by systematic, but free from undue pressure" (p 29-30). The children least ready for systematic reading are typically those whose oral language skills are poorly developed. For these children, ample oral language experiences should be undertaken first. This experience is especially important for children coming from non-English speaking homes. Children also need to learn concepts about written language and its functions—what words and sentences are and can do—so that they have some idea of what adults mean when they talk about reading.

Durkin (1966) has pointed out that for some children ("the-pencil-and-paper kids"), learning to read is a by-product of a desire to write. Writing experiences (not emphasizing good style) are recommended for kindergarten. Simple word processors are already making their way into kindergarten settings.

First Grade Reading—Basal Readers

"The observation that basal programs 'drive' reading instruction is not to be taken lightly. These programs strongly influence how reading is taught in American schools and what students read" (p. 35). It has been estimated that 75 to 90 percent of what goes on in reading is controlled by basal reading programs. Five of them are said to control 75 percent of the market. Early in this century, nearly all students were taught to read through phonic analysis. "...educators such as William S. Gray were responsible for turning American schools away from what they perceived to be the 'heartless drudgery' of the traditional approach. In its place, Gray and others advocated the look-and-say approach. The thinking was that children would make more rapid progress in reading if they identified whole words at a glance, as adults seem to do" (p. 36). Since the mid-fifties, this position has been under attack through such works as Rudolph Flesch's (1955) *Why Johnny Can't Read* and Jeanne Chall's (1967) classic review of the research literature *Learning to Read: The Great Debate.* Chall concluded that the evidence strongly supported the superiority of reading programs that taught phonics as one component of the program. "The picture that emerges from the research is that phonics facilitates word identification and that fast, accurate word identification is a necessary but not sufficient condition for comprehension" (p. 37-37). Most publishers have now incorporated some form of phonics into their programs, but for many it is little more than "window dressing" (Flesch, 1981).

How Should Phonics Be Taught?

The Commission Report spends a considerable space reviewing approaches to teaching phonics. They first note that approaches based on stating rules are counter productive. Students need to be able to say the sounds given the letters, not the rules. Second, they note that many programs teach skills the students can already perform, and they teach unneeded low-frequency combinations. Instruction should focus on the regular letter-to-sound correspondences and the most important irregulars.

There are two major approaches to teaching phonics—the "explicit" and "implicit" approaches. They are apparently not equally effective. The most widely-used reading programs employ implicit phonics. The implicity approach never "says" sounds in isolation. It often uses groups of words having the same sound, such as *sun, sat, sea,* and *so,* to teach a sound. Some children taught with implicit phonics have trouble hearing some sounds like the short /i/ in "sit". "Ironically, therefore, implicit phonics may actually presuppose what it is supposed to teach" (p. 40).

The explicit approach presents sounds in isolation and then the better programs also teach students how to blend the isolated sounds to make words by holding one sound until the next sound begins. For example, *man* is not sounded out "mmm - aaa- nnn", but is sounded out a "mmmaaannn" and then is speeded up to get the word "man." The early need to teach blending skills forces a distinction between *stop sounds* (t, b, k, etc.) and *continuous sounds* (m, a, s, etc.). The later can be held, the former cannot, and therefore need special prompting when used at the beginning of words. "Regrettably, an analysis of published reading programs concluded that several <explicit programs> incorporate procedures for teaching blending that are unlikely to be effective with many children" (p. 39).

Kenneth Goodman (1976) and Frank Smith (1973) have criticized phonetic approach because "they take the child away from meaning." The Commission report points out that such a position is not inherent in phonetic approaches. Some programs have gone too far in pushing phonic instruction—far beyond the point where the student had enough sounds to be reading meaningful words and sentences. "Quite likely the problem is simply a by-product of the *false dichotomy* [emphasis added] between phonics and meaning that has dominated the field of reading for so many years" (p. 42).

Another criticism of many phonics programs is that the words the children are reading have little relation to the phonics they are learning. Thus, they do not get the practice needed for reading to become more automatic.

What works best? "The trend of the data favors explicit phonics" (p. 42). While supporting explicit phonics, the Commission also notes the importance of illustrating sounds by also presenting them in words which serve as exemplars. There is a considerable need to improve the quality of the instructional design in most basal readers on the market today. The Commission is also quite critical of the quality of the stories in basal series in terms of student interest and the information conveyed.

Comprehension in Beginning Reading

The Commission concludes the chapter on Emerging Literacy with a review of evidence on aspects of common reading group practices. The typical lesson has a preparation phase where new words, ideas, and motivating questions can be introduced. Next, comes reading in the group, then discussion, and finally back to a seatwork assignment.

A problem with most preparation phases is that not enough attention is given to developing the *background knowledge* needed to understand the story. "Don't have time" is the excuse. Children remember what they have just read better when background is provided during the preparation stage.

In the reading phase, students take turns reading (usually round robin is better than using volunteers). "There is no substitute for a teacher who reads children good stories. It wets the appetite of children for reading, and provides a model of skillful oral reading. It is a practice that should be continued throughout the grades" (p. 51). As oral reading develops, more silent reading should be planned for. A strong recommendation for improving fluency is to have the students read a passage silently before reading it out loud. Also, reading the same passage several times can improve fluency.

In the discussion phase, comprehension instruction may be provided (although few curricula do), phonics lessons are given, and seatwork is explained. Teachers rely heavily on their manuals in leading discussions. Analysis indicates that many of the questions provided are "too general, leading the children afield; or trivial, focusing their thinking on unimportant details.

"While questions during the preparation and discussion phases of a reading lesson are important, these do not substitute for active, direct instruction. In direct instruction, the teacher explains, models, demonstrates, and illustrates reading skills and strategies that students ought to be using. There is evidence that direct instruction produces gains in reading achievement beyond those that are obtained with less direct means such as questions" (p. 56).

Reading lessons generally should stress "making connections" with what the student understands and appreciation of the content of the story.

Expanding Competency

The Commission believes we need more carefully designed textbooks. Examination of current textbooks show that many fail to "lay bare the fundamental structures of history, geography, health, and science..." (p. 68). Many texts are poorly organized, consisting of little more than "lists of facts loosely related to a theme. Abrupt, unmotivated transitions are frequent. Textbooks are as likely to emphasize a trivial detail or a colorful anecdote as a fundamental principle... When textbooks make clear the connections between motive and action, form and function, or cause and effect, students understands better" (p. 69).

In examining the research on teaching practices that help extend literacy, the Commission recommends *directly teaching* critical concepts and reasoning processes. "*Direct instruction* needs to be distinguished from questioning, discussion, and guided practice. Direct instruction in comprehension means explaining the steps in a thought process that gives birth to comprehension. It may mean that the teacher models a strategy by thinking aloud about how he or she is going about understanding a passage. The instruction includes information on why and when to use the strategy. Instruction of this type is the surest means of developing the strategic processing that was identified earlier as characteristic of skilled readers" (p. 72). The report summarizes a number of studies where directly teaching strategies for attacking text led to the learning of generalized strategies that improved comprehension. The Commission recommends that such strategies be embedded in social science and science lessons and that teachers be given training in how to use these relatively "new" direct approaches to teaching comprehension strategies.

The report next examines activities where students can get independent practice in reading. They are very critical of typical seatwork activities which occupy 70% of the time available for reading, or about an hour per day. The activities often have little value in teaching reading (as when the students are asked to underline the most frequent consonant in a sentence).

The research suggests that the amount of time devoted to workbooks or worksheets is unrelated to year-to-year gains in reading proficiency. In contrast, students average only 7 to 8 minutes a day in silent reading. The amount of time spent on this activity out of school is consistently related to gains in reading achievement. However, a study of fifth graders showed that most students read very little out of school—50% average 4 minutes a day or less, while they average 130 minutes a day watching TV. The research suggests that independent reading may be a major source of vocabulary growth. They conclude that access to good books needs to be improved and silent reading should be encouraged more at both home and school. Children read more books when someone helps to interest them in specific books, when guidance in choosing books is provided, and when time for reading is set aside.

Writing is an activity that supports the development of reading skills. One recent study of elementary school students showed that "only 15% of the school day involved any kind of writing activity. Two-thirds of the writing that did occur was word for word copying in workbooks. Compositions of a paragraph or more in length are infrequent even at the high school level" (p. 80). Students need to be encouraged to write more. "Writing is most beneficial when students have a reason to communicate to a genuine audience" (p. 81).

Effective teachers "schedule reading and writing activities as a priority, move through materials at an appropriate pace, stimulate and sustain children's attention, and arrange for high rates of success" (p. 92). The report examines research on grouping practices and concludes that many current practices slow the progress of lower performing children rather than facilitating it. An improvement in the quality of small group instruction is needed and student should not be "locked into" their reading group for other instruction.

Testing and Reading Instruction

In examining the roles of norm-referenced and criterion-referenced tests in reading programs, the Commission notes that better reading programs tend to use tests more. They suspect that the tests help to motivate teachers to be accountable through the feedback they provide. The report is critical of the use of criterion-referenced tests in mastery-learning type programs where reading is broken down into a series of subgoals. The Commission does not believe that learning to read involves learning one skill, adding another, adding another, etc. Rather, they see learning to read as involving the "close knitting of reading skills that complement and support one another" (p. 97). (My own view is that some mastery approaches have fragmented the learning process, but this need not be the case.) Norm-referenced tests are seen as useful, but may distract from the major goals of teaching reading if instruction just focuses on doing well on the tests. They suggest:

"A more valid assessment of basic reading proficiency than that provided by standardized tests could be obtained by ascertaining whether students can and will do the following: Read aloud unfamiliar selections from grade-appropriate social studies or science textbooks; explain the plots and motivations of the characters in unfamiliar, grade-appropriate fiction; read extensively from books, magazines, and newspapers during leisure time. A simple, practical suggestion is for teachers to tape record the oral reading of each child three times a year and keep the tapes on file for diagnosis and reporting to parents" (p. 99).

Teacher Education and Professional Development

The Commission points to the inadequate time devoted to formal learning and applications in most teacher preparation programs. In elementary education, only about 1/3 of the undergraduate program is devoted to education courses, including "foundation courses". The foundation courses are often criticized as being too theoretical (taught often by teachers who are not familiar with classrooms). In contrast, the practica courses are often seen as too simplistic. At best, two courses are directly related to reading (Reading and Language Arts). The Commission believes teachers need more preparation in reading and other areas and recommends that 5-year training programs be instituted.

On-going professional development efforts often miss the mark. The more successful approaches involve multiple contacts with consultants over a period of time, including visits to classrooms. It also is helpful if a group of teacher band together and give mutual support in learning new strategies. Provisions for assisting the new teacher's entry into the profession should also be undertaken. Experienced, effective teachers might be assigned as mentors during the first year or two.

The closing note before the recommendations focuses on *The Ethos of Effective Schools*. Effective schools have vigorous instructional leadership, usually from a principal. Yet, the report notes that in some states, principals do not even have to know how to teach reading. Effective schools "have high but realistic expectations about the progress that students will make in reading." Effective schools "are characterized by school pride, collegiality, and a sense of community." Effective schools have "order and discipline." Effective schools "maximize the amount of uninterrupted time available for learning" (pp. 113-114).

The Commission finding support the strategies used by DI in teaching reading decoding and comprehension skills. It supports early teaching of reading and getting parents to help in the job. It supports the use of systematic phonics which gets quickly to decoding words, not just sounds. It supports DI's approach to teaching the reasoning skills involved in reading comprehension directly. It supports making the teaching of reading a priority, not to be preempted by outsider interruptions, assemblies, etc. It supports continuous evaluation of reading progress to be sure the approach is work. We do know how to teach reading effectively. We just need to do it.. ◆

References

Anderson, R.C., Hiebert, E.H., Scott, J.A., & Wilkinson, I.A.G. (1985). *Becoming a nation of readers:* The report of the Commission on Reading. Champaign, IL: The Center for the Study of Reading.

Chall, J.S. (1967, 1983). *Learning to read: The great debate.* New York: McGraw-Hill.

Durkin, D. (1966). *Children who read early: Two longitudinal studies.* New York: Teachers College Press

Flesch, R. (1955). *Why Johnny can't read.* New York: Harper.

Flesch, R. (1981). *Why Johnny still can't read.* New York: Harper and Row.

Goodman, K.S. (1976). *The reading process: A psycholinguistic view.* In E.B. Smith, K.S. Goodman, & R. Meredith. *Language and Thinking in School.* New York: Holt, Rinehart & Winston.

Smith, F. (1973). *Decoding: The great fallacy.* In F. Smith (Ed.) *Psycholinguistics and Reading.* New York: Holt, Rinehart, & Winston.

Copies of *Becoming a Nation of Readers* may be ordered from $4.50 each (add $1.00 if overseas) in U.S. funds payable to The University of Illinois—BNR. Send check or money order to Becoming a Nation of Readers, P.O. Box 2774, Station A, Champaign, Illinois, 61820-8774

A Response to the Attack on DI Preschool Programs by Schweinhart, Weikart, & Larner

by Russell Gersten
University of Oregon

From *ADI News* Volume 6, Number 1 (Fall, 1986)

Editor's Note: This article critiques a widely publicized study purporting to attack DISTAR. It should be of considerable interest to ADI members. Reproduced from a publication manuscript with permission from Early Childhood Research Quarterly. *The paper will be published in Vol. 1, 1986, p. 293-302.*

There were several interesting features to the recent exploratory study of the potential long-term effects of preschool by Schweinhart, Weikart, & Larner, (1986) "Consequences of Three Preschool Curriculum Models Through Age 15," which appeared in the first issue of *Early Childhood Research Quarterly*. However, there were also some puzzling aspects to the study, a few curious omissions, two lapses in methodology and several serious flaws in interpretation that need to be pointed out. In particular, the authors' penchant for interpreting non-significant results has led to serious misconceptions about the findings.

This is particularly important because, as most readers of this journal know, the report's findings have been picked up by the popular press, often in a sensationalized oversimplified form. The news media can be almost excused for their hyperbole when researchers fail to follow conventional scientific guidelines. Hechinger (1986) in the *New York Times* reported that "placing children in an early educational pressure cooker can do serious harm," and cautioned against the use of highly structured programs that, according to the study, "appear often to lead to antisocial behavior, delinquency, and even violence later on."

In the study, three groups of 18 fifteen year olds—all having experienced one of three different preschool programs when they were four years old— were compared. In all earlier research reports, the three types of preschool were labeled High Scope, Language, and Child-Centered Nursery. In the current report, the Language group is renamed "Distar." This is an inappropriate choice. *None of the students in the first two waves used the Distar curriculum;* the few students in the third wave who used it, did so for only four months out of the two years of the program. Rather, according to the program directors (Bereiter, 1986: Engelmann, personal communication) these students were taught basic language concepts, shapes, number concepts, colors and letters in a systematic way based on the principles of *Teaching Disadvantaged Children in the Preschool* (Bereiter & Engelmann, 1966). Calling this group *Distar* has created widespread confusion and misinformation about the *Distar* curriculum program which has been successfully used with disadvantaged students in the elementary grades (e.g., Stebbins, et al. 1977).

Due to the extremely small sample size (18 in each experimental preschool condition) and the use of only one site, the authors should have emphasized that this is clearly an *exploratory* study of the later effects of various preschool models. Policy decisions never have been—and hopefully never will be—based on studies involving *brief interviews* and performance on *one test* of a sample of 18 adolescents.

Disparities in Characteristics of the Samples

Consider the characteristics of the three samples involved in the study. They were comparable on many demographic variables, with at least four important exceptions:

1. *Level of mother's education.* The mothers of students in the child-centered Nursery approach had significantly more education. This might tend to bias the results in favor of the Nursery group as Schweinhart, et al., mention.

2. *Unequal representation of the sexes.* Of the 18 students in the Direct Instruction sample, 10 (56%) were boys. For the High Scope sample, 7 (45%) were boys, and for the Nursery School group, 8% were boys. Considering that the focus of the delinquency subscale items is on problems typical of teenage boys, this might tend to bias results in favor of the High Scope and Nursery groups (Bereiter, 1986).

3. *Frequencies of parents out of the home.* Not only did the Direct Instruction group have the highest percentage of single parent families (31 percent as opposed to a mere 12 percent for the Nursery group and 3 percent for High Scope), but more Direct Instruction mothers were employed (44 percent versus 38 and 33 percent respectively). Assuming that the majority of the single parent families were headed by the mother, more of the Direct Instruction students were unsupervised during after school hours, which may have had an impact on their social behavior through adolescence. Here, again there would be a bias in favor of the Nursery group and High Scope.

4. *Amount of preschool experience.* This fourth point is extremely important, and one which is dealt with rather lightly by the authors. Thirty nine percent of the students in the Direct Instruction (DI) and Nursery School samples experienced only one year of preschool. The remaining 61 percent experienced two years of the intervention. In contrast, all of the High Scope students experienced a full two year intervention. This might tend to bias results in favor of High Scope.

In an earlier report on the academic and social progress of these students, Weikart, et al., (1978) argued that students in the first wave should not be included in the analyses. They explained that "since children in Wave 5 (called "the first wave" in the current study) experienced different educational programs as three year olds and were not enrolled for two consecutive years in one of the three CD Project programs they are not included in the longitudinal sample" (p. 19). Since 39 percent of the High Scope sample in the current study experienced one more year of treatment than the DI or Nursery groups, their reasoning still seems sound. Schweinhart, et al., (1986) now assert that the second year of preschool has had no impact on subsequent performance. But how can one be sure it has no impact on the child's social behavior in adolescence until after the evaluation is complete? Clearly, the unequal amounts of preschool is a potential source of bias, favoring the High Scope group.

Thus, despite the random assignment of subjects to treatment at age three or four, several demographic factors were operating then that potentially favor the Nursery and/or High Scope groups. This conclusion is different than Schweinhart, et al.'s, assertion. "We conclude from the analyses presented in Tables 1 and 2 that any outcome differences between the High Scope group and the Distar (sic) group are probably not attributable to group differences in program-entry characteristics" (1986, p. 24).

The Validity of the Measures

The authors used what many would consider a narrow range of measures. Many reasonably objective measures were available, but not utilized: e.g., school achievement (grades and standardized test scores), suspensions and retentions, truancy, and special education placements. Unlike previous work by Weikart and colleagues, this study relies almost exclusively on self-report measure, all of uncertain reliability and unknown, rather dubious validity.

The self-report measures included: (a) Perceived

Locus of Control the Bialer (1961) scale, with an unacceptably low reliability (coefficient alpha of .34); and (b) a measure of self-esteem, the Rosenberg (1965) scale, with an acceptable coefficient alpha reliability of .70. (No validity data reported.) The crux of the evaluation hinged on data gathered by self-report procedures, by a structured interview covering a range of *antisocial activities* which was rather unfortunately labeled "Juvenile Delinquency Scale," and by a rather scattered series of questions on family life and life at school.

The only reliabilities reported are internal consistency estimates. While coefficient alpha is an acceptable way to gauge reliability of academic measures, it is inappropriate for self-report measures, since if an individual is dishonest or distorts information in either direction—either by concealing antisocial or shameful criminal activities or by "boasting" about non-existent criminal activities—he or she would actually inflate the coefficient alpha. A measure of temporal stability would have been superior.

No validity data are provided on these self-report measures and, reading the results of some items on the scale, one wonders about the teenagers' candor. When asked, "Have you ever . . . . argued or fought with parents?" the mean score for the High Scope students was 1.11 (where 0 = never, 1 or 2 means once or twice in your life, 3 means three or four times, and 4 means 5 or more times). It seems amazing that any adolescent, let alone those adolescents, half of whom had been arrested by the age of 15, with an average of 2.2 suspensions from school, had argued or fought with their parents only once in their lifetimes.

The other measures included a paper and pencil multiple choice test on knowledge deemed necessary "for educational and economic success in modern society," called the Adult Performance Level Survey (APL), and information on number of suspensions gathered from school officials. The rationale for administering the APL was that it "can provide insights into the real-world competence these adolescents have developed in applying skills learned in school to the demands of adult life" (p. 28). Reliabilities for each subscale are reported from .32 to .63 with a median of .58. The reliability of this measure borders on being unacceptably low. No validity information is reported. Information on the number of suspensions was collected, but, for some strange reason, not analyzed by curriculum model. Only the overall mean was reported.

Interpretation of Results

It is interesting to speculate what would have happened if the data from this study had been reported and analyzed by a team of independent evaluators, such as the professional groups used to analyze and discuss the evaluation of Follow Through—rather than a team of researchers directly affiliated with one of the three curriculum approaches evaluated. Unlike Schweinhart, et al., (1986) they would use the conventional .05 level of significance. The authors would let the reader know which statistical tests were used, and perform post-hoc tests to delineate which of the three groups were significantly different from each other. Considering the heavily skewed distribution on most of the items dealing with antisocial behavior, they would need to use nonparametric tests. They would need to deal with the Delinquency Scale on an item-by-item basis, rather than creating five subscales out of a mere 18 items. Of course, the evaluators would not interpret non-significant findings. And thus, I believe a quite different picture would emerge.

This mythical report would begin with the objective data—a mean of 2.2 suspensions for the entire sample of 54. It would indicate that no significant differences were found among the three samples on this measure, which is presumably the case. Next, the report would indicate that half of the students in the total sample reported they had been arrested at least once by age of 15. Again, apparently no significant differences existed between the groups.

The report would next indicate that there were no significant differences in self-esteem, as measured by the Rosenberg scale. (The locus of control measure would be dropped due to its low reliability.) Table 1 indicates the only items for which significant differences were found. This of course assumes the results for these items would still be significant when appropriate (nonparametric) statistical tests were used to compare differences.

The evaluators would indicate that the lack of significant differences between the three preschool models is unsurprising, considering the array of experiences in school and out of school since the age of four that were much more likely to have an impact on their lives. The report would conclude that, while preschool appeared to help all three groups during the primary grades (as evidenced by some elevation in IQ scores, and reasonable achievement scores in first and second grade), the students don't appear to be doing too well in junior high school; many have serious problems in school and with the courts. They might well conclude that more effective education in the elementary schools and junior high schools might have helped.

Summary of Findings

What can one conclude from these data? First, the students, overall, are not doing very well; half have been arrested, and many have been suspended from school. For suspension rate and self-reports of arrests, no differences were found between the three curricu-

Table 1. Mean Scores on Items Where Significant Differences Were Found by Schweinhart, et. al., (1968)[a]

	Direct Instruction	High Scope	Nursery	p^c
Adult Performance Level Survey				
Occupational Knowledge	2.4	3.7	3.7	.04
(All other scales non-significant)				
Self Report				
1. Have you ever run away from home?[b]	.38	.17	0	.02
2 Appointed to a school office or job?	0	12%	33%	.02
3. Participation in sports.				.02
Often	17%	50%	44%	
Sometimes	28%	44%	28%	
Never	56%	6%	6%	
4. How does your family feel about how you're doing?				.03
Great	0	6%	6%	
Alright	67%	94%	89%	
Poorly	33%	0%	6%	

[a] Standard deviations unavailable
[b] 0=never, 1=once, 2=twice, 3=three or four times, 4=5 or more
[c] Type of statistical test performed is unavailable

lum models. Nor were there differences in self-esteem. Of the numerous self-report categories (including damage to school property, serious fights, stealing, use of weapons) no differences are significant.

The achievement level of the students is unknown. No normative data are presented for the Adult Performance Level survey, so we cannot ascertain how these students fared compared to their peers. All we know is that of the ten subtests of the Adult Performance Scale, differences between the three samples were significant on only one.

Significant differences were found in only three areas—sports participation, being appointed to an office or job at school, and running away from home. Though many would consider playing basketball or being on the track team nice, surely failure to do so is not a cause for alarm. It is unclear that any of these are evidence of hardcore juvenile delinquency. Students may well have good reasons for running away from home.

There are many ways to approach item 4 on Table 1. One third of the DI students thought that their family felt they were doing poorly, while none of the High Scope students, and only one of the other 18 Nursery students felt this was the case. Considering the arrest records and frequency of suspensions, which group presents a more *accurate* picture? This mythical report would conclude that of the many, many comparisons made on the self-report measures, only four were significant—and none were in crucial areas. None had anything to do with delinquency. These few differences could be due to the somewhat different male/female ratios between the groups or perhaps even unequal exposure to preschool experience.

Achievement and IQ During the Elementary Grades

In their report, Schweinhart, et al., (1986) also devote a considerable amount of time to summarizing their earlier findings on the achievement of these students in first and second grades, and their IQ's up through fourth grade. Here, too, some of their interpretations are a bit misleading. The mean growth in IQ between the ages of 4 and 10 for all three groups is impressive. Students tended to begin the program with a mean IQ of 79 at the age of 3, and the average IQ at age 10 was 93. However, the readers should note that the sample size was 55 at pretest and a mere 29 at posttest. Schweinhart, et al., might also have pointed out that the predictive validity of IQ scores obtained before the age of 5 is close to zero (McCall et al., 1975). The reason for this may be clear if one thinks about the difference between the type of items a three year old takes versus those a ten year old takes on the Stanford-Binet.

Though there were not significant differences in IQ scores at age 10 between the three experimental samples, the Direct Instruction group mean was over one-third of a standard deviation higher. It is odd that this is the one time Schweinhart, et al., *Chose not to interpret a non-significant finding*, one that meets common criteria of educational significance. It is also strange that no IQ measures were administered to these 15 year olds.

The achievement data based on the California Achievement Test indicates that no significant differences appeared between the three samples in achievement in first and second grade. However, the achievement level at both grades was at or near grade level. Again, one wonders why no measures of achievement were collected or reported for students at age 15. The high suspension rate in junior high school may indicate that academic achievement is not at a very high level. Other longitudinal studies of low income, minority students have noted increasing losses against the norm sample in the later elementary grades and in middle school (Becker & Gersten, 1982; Stebbins, et al., 1977).

Summary of Problems

Aware of the problems in extrapolating from a study based on such a small sample, Schweinhart, et al., (1986) indicated "this report requires *major restraint in its use and interpretation*" (emphasis added), (p. 43). Yet in the next sequence, the policy implications are clearly drawn. Their choice of the phrase "pressure cooker" to describe the Bereiter-Engelmann approach for teaching language concepts was immediately picked up by the popular press. This is despite the fact that a naturalistic observational study failed to find any significant differences, in how teachers and students interacted, between the alleged "pressure cooker" approach and Weikart's own "cognitive" approach (Seifert, 1969). The percent of verbal feedback was essentially the same for the two groups, as was the percent of time teachers spent on management. Interestingly, neither were there any differences between the two approaches in the percent of time teachers spent on affect issues. The only difference was that significantly more verbal interaction went on in the Direct Instruction preschool.

In a response to the *New York Times* article (Hechinger, 1986), after noting the unequal demographics, Carl Bereiter (1986) stated, "For those who associate direct instruction with harsh discipline, it may be important to know that the supervisor of [the] direct instruction group reported (McClelland, 1970) that punishment was not used and discipline problems were virtually nonexistent."

Bereiter then asked, "How could direct instruction at age three of four have led to delinquency at age 15?" It is equally reasonable to assume that the unequal demographics, the high proportion of males in the direct instruction sample, and/or the higher number of students coming from homes without parent supervision, contributed to the few significant differences found. In addition, the children's experiences in kindergarten, elementary school and junior high school would certainly have some impact on their lives at age 15. Yet nothing has been recorded about the children's later educational experiences. Material on the current demographics and status of the students' families might also help understand some of the differences. Obviously, home situations change over a 12-year period, particularly in a high unemployment state such as Michigan, and these factors should have been recorded.

Though a few statements appeared in the article formally stating that further research is needed before firm policy conclusions can be drawn, the authors make numerous inferences regarding the impact of the curricula used in preschool on children's future delinquency. At times, the text is written as if self-reports were the same as actual behavior (e.g., "The Distar group engaged in five times as many acts of property violence. . ." (p. 34)). The authors' setting of an extremely liberal .10 significance level is inappropriate. In a study such as this, we need to be sure *before* inferences are made.

The media, of course, picked up on the findings without the reservations occasionally expressed by the authors. Titles such as "Preschool Pressure, Later Difficulties Linked in Study" from the April 23, 1986 *Education Week* give a sense of the typical thrust of the media interpretations. The *New York Times'* reporting that "direct instruction in preschool leads to twice the amount of delinquency (Hechinger, 1986) is typical. The presentation style of the original article led to these misinterpretations. In fact, students from the three groups were not significantly different on more than five measures. Some of the areas of difference—sports participation, being appointed to a school office—were not of a dramatic social consequence. One group of 18 children seemed to get along more poorly with their families than the other two groups as evidenced by one of the interview items and self-reports of running away from home. Whether this is due to the academic emphasis of the preschool is dubious.

Finally, there is a need to point out that the situation is far less sanguine for these students—regardless of type of preschool program—than Schweinhart, et al., admit. The glaring omission of data on achievement and school attendance, and the failure to fully analyze the data on suspensions and arrest records are curious. All we know is that the students were doing well in school at the end of the second grade. The meager evidence presented here suggests major problems by the ninth grade for all three groups of students.

It appears that something more than a special preschool program is needed to make a difference in these children's lives. Superior elementary school programs are a necessity. Here, the overwhelming consensus of multi-site, large scale independently conducted research studies is that approaches that use some form of direct instruction would lead to superior academic growth (Stallings, 1975; Stebbins, et al., 1977). In addition, some of our more recent large scale multi-site research, with samples of approximately 1000 (Becker & Gersten, 1982; Gersten, Carnine & Keating, 1984; year, 1984), show enduring effects for this approach—including a reduction in dropout rate and increased college acceptance. ◆

Author's note. The author wishes to thank Thomas Keating and Gerald Tindal for their feedback on earlier drafts of this manuscript, and Sue Cox for her assistance in preparing the various drafts of the paper.

References

Becker, W.C., & Gersten, R., (1982) A follow-up of Follow Through: Meta-analysis of the later effects of the Direct Instruction Model. *American Educational Research Journal, 19*, 75-93.

Bereiter, C. (1986). "Mountains of evidence," said to contradict study on effects of preschool. *Education Week, 5* (37), 19.

Bereiter, C. & Engelmann, S. (1966). *Teaching the disadvantaged child in the preschool.* Englewood Cliff, N.J.: Prentice-Hall.

Bailer, I. (1961) Conceptualization of success and failure in mentally retarded and normal children. *Journal on Personality*, 29, 301-333.

Gersten, R., Carnine, D., & Keating, T., (1984, April). The lasting impact of the Direct Instruction Follow Through Program: A Longitudinal study of 1500 students. Paper presented at American Educational Research Association, New Orleans, LA.

Hechinger, F.M. (1986, April, 22) Preschool programs. *New York Times.*

McCall, R.B., Appelbaum, M.I., & Hogarty, P.W., (1973). *Developmental changes in mental performance.* Monographs of the Society for Research in Child Development, 38 (3).

McClelland, D., Hiatt, L., Mainwaring, S., & Weathers, T., (1970). *The language training curriculum.* Ypsilanti, MI: High/Scope Educational Research Foundation.

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

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, N.J.: Princeton University Press.

Schweinhart, L.J., & Weikart, D.P., (1980). Young children grow up: The effects of the Perry Preschool program on youths through age 15 (Monograph of the High/ Scope Educational Research Foundation, 7). Ypsilanti, MI: High/Scope Press.

Schweinhart, L.J., Weikart, D.P. & Larner, M.B. (1986). Consequences of three preschool curriculum models through age 15. *Early Childhood Research Quarterly, 1*, 15-45.

Seifert, K. (1969) Comparison of verbal interaction in two preschool programs. *Young Children, 24(9)*, 350-355.

Stallings, J., (1975). Implementation and child effects of teaching practices in Follow Through classrooms. *Monographs of the Society for Research in Child Development*, 40.

Stebbins, L., St. Pierre, R.G., Proper, E.C., Anderson, R.B. & Cerva T.R., (1977). *Education as experimentation: A Planned variation Model, Volumes IV A-D.* Cambridge, Mass: ABT Associates.

Weikart, D.P., Epstein, A.S., Schweinhart, L. & Bond, J.T. (1978). *The Ypsilanti preschool curriculum demonstration project: Preschool years and longitudinal results*, Ypsilanti, MI: High/Scope Educational Research Foundation.