DOCUMENT RESUME

ED 228 320 TM 830 246

AUTHOR Pecheone, Raymond L.; Nearine, Robert J.
TITLE An Exploratory Examination of Sustaining Program Effects Using Multiple Outcome Measures.
PUB DATE Apr 83
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Attribution Theory; *Compensatory Education; Educational Policy; Elementary Education; Miscue Analysis; Pretests Posttests; *Program Effectiveness; *Remedial Reading; *School Districts; *State Departments of Education; Validated Programs
IDENTIFIERS *Elementary Secondary Education Act Title I; *Intensive Reading Instructional Team; Metropolitan Achievement Tests; Stanford Diagnostic Reading Test

ABSTRACT

This study examines the long term effects of a compensatory reading project, the Intensive Reading Instructional Team (IRIT), which has been validated both by the Education Department's Joint Dissemination and Review Panel and the Connecticut State Department of Education. A second purpose is to demonstrate to both the hosting district and to other Connecticut school districts that a sustaining effects study could effectively impact on local policy making. This demonstration provides practical examples of how a Title I project could comply with long term study requirements and produce useful information for local policy makers as well. Finally, this study serves as an example of the type of partnership which can be established between a school district and state researchers. The following measures were used in the multiple assessment strategies: (1) Metropolitan Achievement Test, (2) Stanford Diagnostic Reading Test, (3) Miscue Analysis, (4) two attribution measures of pupil's perceptions of their confidence and competence in reading, and (5) attendance data. Although the sustaining effects study is not completed, several policy-based activities have begun, including the use of the IRIT sustaining effects model as a base for generating other long-term program data. (Author/PN)

AN EXPLORATORY EXAMINATION OF SUSTAINING PROGRAM EFFECTS USING MULTIPLE OUTCOME MEASURES

Raymond L. Pecheone
Connecticut State Department of Education

&

Robert J. Nearine
Hartford, Connecticut Public Schools

Introduction

The purposes of this investigation were threefold. First, this study was designed to examine the long term effects of a compensatory reading project which had been validated both by the Education Department's Joint Dissemination and Review Panel (JDRP)[1] and the Connecticut State Department of Education.[2]

A second and equally important purpose was to demonstrate to both the hosting district and to other Connecticut school districts that a sustaining effects study could effectively impact on local policy making. This demonstration would provide practical examples of how a Title I (Chapter I) project could comply with long term study requirements and produce useful information for local policy makers as well.[3]

Finally, this study serves as an example of the type of partnership which can be established between a school district and State researchers thus supporting the contention that appropriate State technical assistance is but a telephone call away.

Study information is expected to impact on four populations: project staff, school board members, district policy makers, and out of district Title I (Chapter I) and compensatory program administrators.

The Program

The Intensive Reading Instructional Team (IRIT) program is a unique approach to remediation.[4] Designed by Hartford staff in early 1964 as an individualized approach to reading, the IRIT program serves about 700 youngsters each year. IRIT's are characterized by low pupil teacher ratios and the use of a wide variety of materials and equipment. In Hartford, forty-five pupils per team are selected for each of the three 10-week instructional cycles and are instructed by the three team teachers who specialize in the reading areas of encoding and decoding, individualized reading, and vocabulary/comprehension. Within each cycle, smaller classes of approximately 15

Presented at the annual meeting of the American Educational Research Association, Montreal, P.Q., April 1983.

youngsters move from one area to the next at approximately one hour intervals each morning for reading instruction. In the afternoon, these pupils return to their classrooms for instruction in other subject areas. IRIT teachers work with the same children for the entire morning and use the afternoons for preparing individualized lessons, coordinating the instructional approaches for each of the 45 students, updating pupil records, and meeting with teachers and parents from the sending schools. Therefore it is possible for staff to provide intensive instruction to students and a high degree of motivation as well. Further, since each team typically selects cycle enrollees from no more than three schools and ten classrooms, the result is a lower sending classroom teacher ratio with more morning time available to work on reading and language arts for the youngsters who remain behind.

Methodology

Sample

All 1981-82 school year IRIT student enrollees in grades three through six participated in this study for a total of 698 students. This IRIT enrollment included 241 third graders, 212 fourth graders, 225 fifth graders, and 20 sixth graders. Because the IRIT is a Title I (Chapter I) funded program, students were selected for enrollment on the basis of their performance on a nationally normed standardized test. Hartford uses the 23rd percentile to identify the students who would be eligible for IRIT services. However, many more students are eligible than the program is able to serve. From this pool of non-served students a comparison group was identified by matching students on the basis of their prior year's MAT reading score. The comparison group was made up of 24 students in grade three, 38 students in grade four, 33 students in grade five, and 29 students in grade six, for a total of 124 students. These children received no other compensatory services during the 1981-82 school year.

Research Questions

The investigation was designed to address the following research questions:

1. What are the short term and long term effects of IRIT instruction after pupils return to the "regular" classroom?
2. How does the method of the reading assessment effect the interpretation of the results using multiple assessment strategies: i.e. Metropolitan Achievement Test (MAT), Stanford Diagnostic Reading Test (SDRT), and a Miscue Analysis (MA)?
3. How does a pupil's perceptions of their Confidence in Reading and Competence in Reading, two attribution measures which were developed for the project, relate to achievement performance over time?

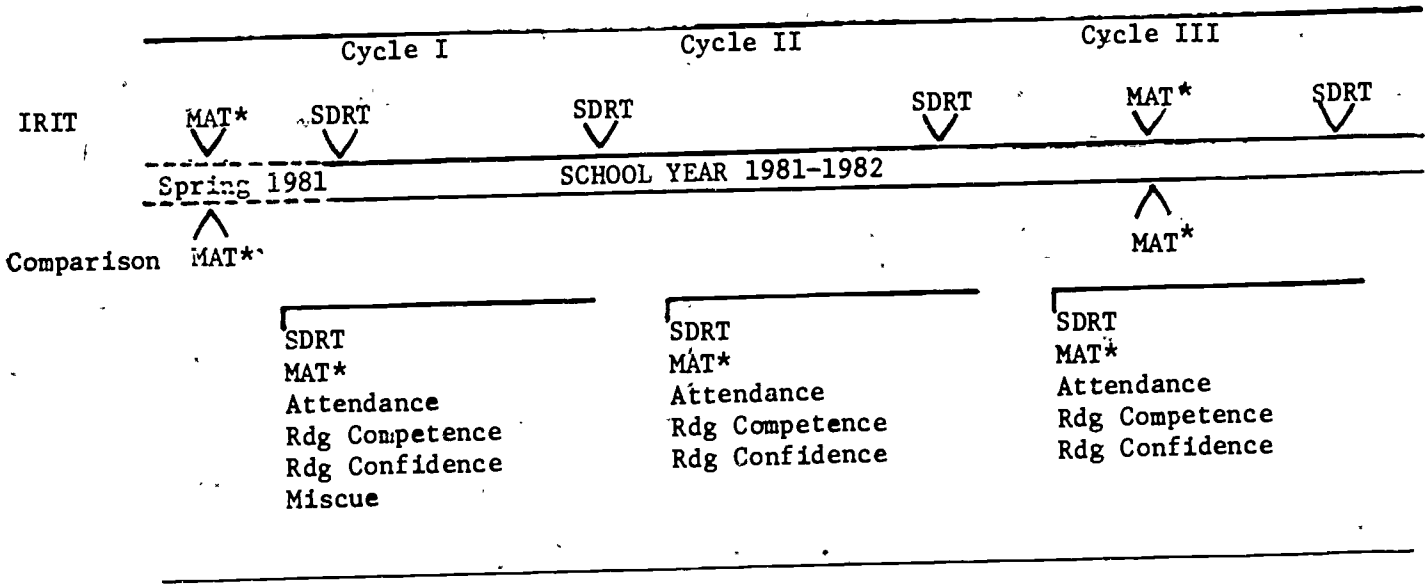
In addition to the more formal research questions, the investigation was also intended to provide documentation of the sustaining effects study process which could be applied to other compensatory programs. Here the intent was to help districts become aware of the policy-making capabilities which information from such a study could provide.

Research Design

The IRIT program is divided into three 10-week cycles each year. This structure is shown in Figure 1, as are the following data sources which were included in the study: Metropolitan Achievement Test (MAT), Stanford Diagnostic Reading Test (SDRT), Miscue Analysis (MA), two attribution measures (Confidence in Reading and Competence in Reading), and attendance data.

FIGURE 1

IRIT Study Design**



*District wide testing administered in late fall of each year. Subsequent andlysis will include spring 1983 MAT test data.

**Individualized Reading Inventories (IRI) and the Metropolitan Instructional Test (MIT) are routinely administered on an incycle and fall basis. These instruments provide data which may be examined in subsequent analyses.

Figure 1 also shows that a between cycle comparison of IRIT students was possible. Pupils who had received IRIT services were compared with those pupils who were entering an IRIT cycle. For example, post cycle I pupils were compared with entering cycle II and post cycle II was compared with entering cycle III. To accomplish this between cycle assessment, an attempt was made to balance the distribution of students within each grade level, grades 3-6, across the three IRIT cycles.

At least 5 data points are available to assess long term effects within the IRIT project. These include district wide spring 1981, spring 1982, fall 1982 MAT testing, and SDRT pre and post tests within each of the three instructional cycles. In addition, a comparison group which was receiving no additional compensatory reading instruction was used to establish baseline information as well as to monitor growth in reading over time.

Miscue Design

Miscue analysis is a holistic reading technique which is individually administered. Through this process a 4 x 4 matrix of scores which assesses a student's oral reading and reading comprehension abilities is generated. Taped oral reading selections from designated basal texts were individually administered on a pull-out basis by locally trained reading specialists on a fall 1981 pre test and a spring 1982 post test basis. From the IRIT program, 75 cycle I and II pupils were randomly selected to participate in the oral reading while 45 students were randomly selected from the comparison group. Every attempt was made not to label these students as being either a member of the IRIT or the comparison group.

The use of the miscue analysis provided an opportunity to examine reading comprehension from a performance based perspective. Since this examination focuses on a students' ability to read orally and with comprehension, the holistic assessment of reading comprehension is different from assessments which use a norm referenced standardized reading test. This analysis enables the researcher to have a better understanding of student reading achievement and provides and opportunity to examine the relationship between two reading assessment strategies.

Analysis Strategies

The comparability of the comparison group(s) with IRIT cycles, I, II, and III students was tested using an ANOVA procedure to determine whether these groups were random "in effect". Based on these results scores can be adjusted, when appropriate, using either a covarience or principal axis procedure to equate the groups.

The analysis of the outcome data from multiple data sources was conducted using a Lindquist Type Repeated Measures design (I and III) to assess long term effects (Question 1) while a correlational analysis was employed to assess the relationships among the outcome measures (Questions 2 and 3).[5]

Two types of scales were used to analyze the outcome measures. Standard scores were used with the MAT and SDRT while scale scores were generated for the affective instruments. Standard scores were converted to normal curve equivalents (NCE) and were averaged to place the standardized achievement data on a common scale. Scale scores were generated for both affective measures by summing across the item values to obtain a total score. Since reliability and validity studies on the affective measures are still being conducted, these data should be interpreted cautiously.

Findings

Pretest/Posttest Comparisons

Table 1 presents the results of comparisons of the NCE means for the IRIT and comparison groups on the following measures: MAT reading, MAT language, SDRT reading, and miscue analysis. The standardized test scores which are included in the table represent those students who scored less than 34 NCE (-23 %ile) on the on-level reading pretest which was administered in spring, 1981.

TABLE 1

Related T-Test Across
Outcome Measures By Group

Groups / Variables	N	Mean Pretest SS	Mean Pretest NCE	Mean Posttest SS	Mean Posttest NCE	Gain SS	Gain NCE	t-test
Cycle I								
MAT-Rdg	165	573	27	626	34	+53	+ 7	-6.45*
MAT-Lang	166	455	29	548	38	+93	+ 9	-8.19*
SDRT-Rdg	256	391	28	423	40	+32	+12	-15.89*
Cycle II								
MAT-Rdg	131	599	28	646	34	+47	+ 6	-6.54*
MAT-Lang	128	505	31	583	38	+78	+ 7	-6.29*
SDRT-Rdg	215	413	32	441	42	+28	+10	-14.67*
Miscue	16	457	-	493	-	+35	-	-13.48*
Cycle III								
MAT-Rdg	132	594	28	642	36	+48	+ 8	-13.44*
MAT-Lang	129	493	31	597	44	+104	+13	-12.61*
SDRT-Rdg	216	405	38	436	48	+31	+10	-13.56*
Comparison								
MAT-Rdg	70	616	26	658	34	+42	+ 8	-5.91*
MAT-Lang	69	577	37	644	44	+67	+ 7	-5.22*
Miscue	26	594	-	609	-	+15	-	9.86*

* Significant $p < .05$

For IRIT and comparison group students, NCE gains from pretest to posttest were obtained on the MAT reading and language subtests. NCE gains for IRIT students in reading ranged from 6 to 9 NCEs and language arts gains ranged from 7 to 13 NCEs. The comparison group also demonstrated gains in reading (8 NCEs) and in language arts (7 NCEs). Title I guidelines suggest that a gain of 5 or more NCEs is generally considered educationally meaningful. Means and standard deviations of pre and post IRIT training were calculated and differences between means were analyzed by use of a correlated t-test. Statistically significant gains were found for all 3 IRIT cycles as well as for the comparison group.

Correlated t-tests were also used to assess mean gains on the SDRT within cycles I, II, and III. NCE gains within each 10-week cycle ranged from 10 NCEs (cycles II and III) to 12 NCEs (cycle I). These mean gains were found to be statistically significant within each of the three cycles.

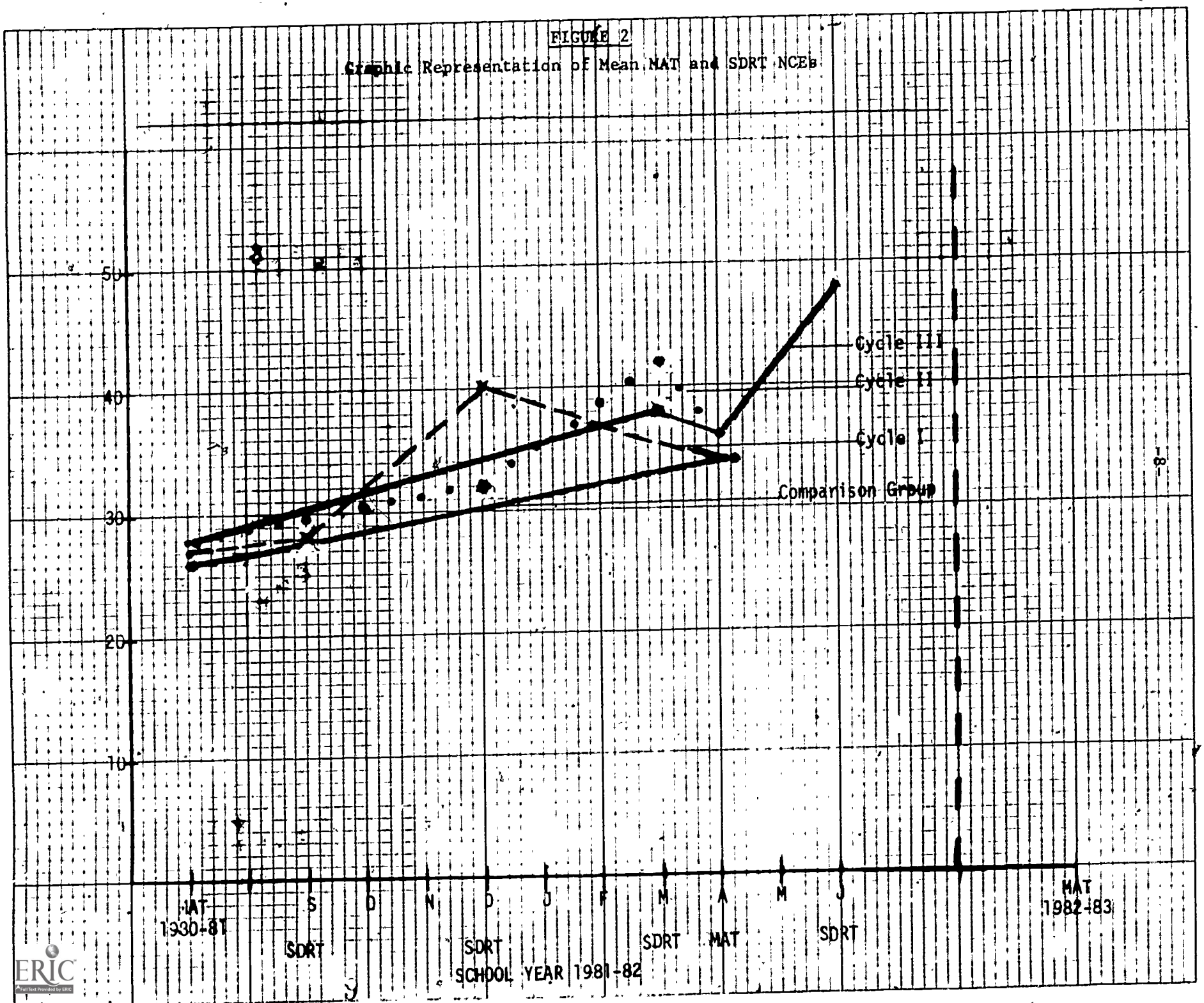
Finally, the miscue analysis assessment strategy was analyzed on a pretest to posttest basis, primarily for cycle II and the comparison group. Standard score mean gains were found for both cycle II students (+35) and the comparison group (+15). Statistically significant overall mean gains ($p < .05$) were also found on the miscue analysis for both the IRIT and comparison group students.

Graphic Display of Reading Performance

Figure 2 displays the mean reading performance of the students in each IRIT cycle and the comparison group on the following outcome measures: MAT reading pretest, MAT reading posttest, SDRT pretest, SDRT posttest. The SDRT was administered only to the IRIT students. NCEs were used to graph MAT and SDRT results in order to use a common metric across both types of reading assessments.

FIGURE 2

Graphic Representation of Mean MAT and SDRT NCEs



An examination of Figure 2 shows two sets of comparisons. Students in the IRIT cycle(s) are compared with the comparison group on the MAT reading comprehension pretest and postest, while reading comprehension performance for students exiting an IRIT cycle is compared with performance of students entering the next cycle. Here, post cycle I students are compared with entering cycle II students and post cycle II students are compared with students entering cycle III as measured by the SDRT.

Further inspection of Figure 2 reveals that the mean MAT reading performance NCE scores on both the pretest and posttest were essentially the same. The scores themselves can be found in Table 1.

A one-way analysis of variance was performed across both MAT testing points to assess whether mean NCE differences were statistically significant across groups. This analysis showed no significant differences between groups on the MAT reading comprehension pretest ($F = .84$, $df = 3/475$, $p > .05$). On the posttest analysis, again no significant differences were found on the MAT reading subtest ($F = .96$, $df = 2/586$, $p > .05$). These findings suggest that the researchers attempt to balance group selection for each cycle and for the comparison group was successful thus indicating that the groups could be considered random "in effect".

IRIT student reading performance for the 1981-82 school year testing points are also shown in Figure 2. Students who completed an IRIT cycle performed higher than entering students in the next IRIT cycle. Here, the mean reading comprehension score for exiting cycle I students (December) was 40 NCEs as compared with 32 NCEs for the students who were entering cycle II. For the exiting cycle II students, the mean reading comprehension score (March) was 42 NCEs as compared with 38 NCEs for the students who were entering cycle III.

A t-test was used to compare cycle I students with cycle II students and cycle II students with students in cycle III. In the first comparison there was a significant increase in reading performance which favored cycle I students as compared to cycle II students ($t = 9.79$, $df = 294$, $p < .05$), while in the second comparison a significant increase in reading performance for cycle II students when compared to students in cycle III ($t = 6.53$, $df = 262$, $p < .05$) was found. These findings indicate that participation in the IRIT program had a significant impact on student reading performance. Significant reading gains were found from pretest to posttest on both the MAT and the SDRT within each IRIT cycle. This finding suggests that IRIT students sustained their gains in reading throughout the 1981-82 school year.

These students will be followed up in future studies to assess reading achievement over additional periods of time.

Assessing Reading Impact Across Four Data Points

To test differences between the three IRIT cycles, an analysis of varience using a three factor design with repeated measures on one factor was employed. The measure of performance in this analysis was the students' pretest and posttest reading scores on both the MAT and the SDRT. The two blocking factors in this design were IRIT cycles and grade over the four testing points. These data were treated separately by a Type III analyses of varience (Lindquist) and are presented in Table 2.

TABLE 2

Analysis of Varience for Reading Test Scores

Across IRIT Cycles and Grades

Source	Df	Ms	F-Ratio
Between			
Cycle	2	37326.6	3.80*
Grade	2	17958.5	1.82
Cycle x Grade	4	15398.2	1.56
Error Between	394	9800.2	Not tested
Within			
Testing	3	79676.1	158.30**
Cycle x Testing	6	6696.9	1.33
Grade x Testing	6	8976.7	1.78
Cycle x Grade x Testing	12	7525.8	1.50
Error within	1182	5032.0	Not tested

* $p < .05$

** $p < .001$

Two significant effects were obtained from the analysis of these testing data: Cycle ($F = 3.80$), and testing ($F = 158.3$). None of the interaction factors were statistically significant.

It seems, therefore, that growth in reading achievement did occur across the four testing points. This finding is further reinforced by the univariate analyzes presented in Table 1. In addition, it also appears that students in one of the cycles seemed to perform better on the reading achivement measures than students in other cycles. This main effect (cycle) tests the aggregate test performance for each cycle across the four testing points. Appropriate post hoc analyses will be performed to identify which cycles produced the main effect.

Intercorrelations Among Outcome Measures

Table 3 shows Pearson product moment correlations which were obtained across eight outcome measures. The Table also shows a predictable pattern with reasonably high correlations between the reading achievement measures and low correlations between the affective and achievement measures. Although this pattern is not surprising, one relationship should be further highlighted. The miscue analysis score shows low correlations with both the MAT ($r = .07$) and the SDRT ($r = .23$). These low correlations suggest that the miscue is measuring a dimension of reading that is different from that measured by standardized achievement tests. Since the miscue analysis was included in the study based on this assumption, the finding reinforces the contention that a performance based assessment may help to better understand the impact of the IRIT. Further examinations of the relationship between miscue analysis and standardized test scores will be carried out at a later date.

TABLE 3

Pearson Product Moment Correlations
Across Selected Outcome Measures

Variable	1	2	3	4	5	6	7	8
1 MAT-Pre Rdg	1.00	.62	.57	.60	.23	.13	-.11	-.06
2 MAT-Post Rdg		1.00	.65	.70	.11	.15	-.005	+.03
3 SDRT-Pre Rdg			1.00	.79	-.11	-.07	-.004	-.01
4 SDRT-Post Rdg				1.00	.07	.17	-.03	-.01
5 Miscue-Pre					1.00	.73	-.24	-.15
6 Miscue-Post						1.00	-.11	.13
7 Reading Confidence							1.00	.70
8 Reading Competence								1.00

Group Comparisons on the Affective Measures

Table 4 contains the results of pretest and posttest affective comparisons using the (1) Confidence in Reading and, (2) Competence In Reading. These instruments were developed specifically for this project with validation studies presently being conducted.

TABLE 4

Mean Performance on Affective Measures

Groups / Variables	N	Pretest Mean SS	Posttest Mean SS	Gain	Correlated t-test
Cycle I					
Confidence	63	14.47	14.11	-.36	1.41
Competence#	61	44.87	46.20	1.33	-.30
Cycle II					
Confidence	176	14.58	14.52	-.06	.40
Competence#	175	54.82	47.10	-7.72	3.09*
Cycle III					
Confidence	169	14.51	14.36	-.15	.96
Competence#	165	56.40	50.56	-5.84	2.16*
Comparison					
Confidence	35	14.51	14.58	-.03	.10
Competence#	36	73.27	72.94	-.33	.06

Since the reading competence scale is reversed, negative gains represent an increase in student perceptions of their competence to read.

* $p < .05$

Table 4 shows that small differences were found on both affective measures from pretest to posttest. The reading competence measure for cycles II and III produced the only statistically significant findings. This measure assessed students' perceptions of their ability (competence) to engage in a variety of reading activities and served as an estimate of how much effort they perceived was put into reading activities. These findings suggest that students in IRIT feel that they have increased their competence to read. While this is a very positive finding, since the testing points covered only the 10-week IRIT treatment cycle, more information is needed on whether these perceptions are sustained over time. A further study of long term effects on the affective measures will be conducted in the spring, 1983.

Discussion

Preliminary Findings

The analysis of data collected from five points which extended from spring 1981 through the 1981-82 school year provided some insight into the long term effects of the IRIT program. This insight is expected to be supported by data which will be collected in the spring of the current 1983 school year. Analyses completed to date suggest that:

1. The impact of the IRIT is not determined by the time of year during which students enter the program. As a group cycles I, II, and III students performed at about the same level on the MAT and SDRT standardized achievement measures.
2. Within cycle comparisons seemed to demonstrate clearly that students gained significantly in their reading achievement as a result of IRIT program participation. This finding is further substantiated by the use of "entering cycles" as comparison groups. In both instances where post cycle I students were compared with entering cycle II and post cycle II students were compared with students who were entering cycle III, significant differences were found favoring the cycle which had received the IRIT treatment. Post treatment cycles performed significantly better than their "non-treatement" student peers.
3. While data suggest that cycle I students do not lose ground in reading over the course of the school year, it is difficult to predict longer term IRIT program effects on the basis of current study data. In consequence, two additional data points were established and

plans have been made to collect MAT data in April 1983 and in April of the following 1984 school year. These data will be analyzed to assess longer term program effects.

4. Finally, the impact of the IRIT program on reading competence and confidence and the holistic assessment of reading need further study. Again, further affective and miscue analysis inquiries are projected for the coming year.

Although the sustaining effects study has not as yet been completed several policy based activities have already been initiated. A synopsis of the current paper is presently being written for distribution to staff, conferencing with study participants has been scheduled, and it is expected that preliminary study results will be presented to the district Board. At the same time, the IRIT sustaining effects model is being used as a base for generating other long term program data. And since these data are being collected from high priority remedial programs, it is expected that information will be available for the first time (1) to help district policy makers determine the differential impacts of programs on students over time and (2), to provide these staffs with a rational means for projecting remedial efforts over a five-year period. This data based planning model is in sharp contrast to past planning efforts which were generally crises oriented, and were responding to a given funding availability. The overall impact of this plan, of course, is yet to be seen.

Still another planning activity is being undertaken. The office of Evaluation, Research, & Testing will prepare some vignettes which will describe sustaining effects activities. Here the intent is to produce brief, easily read documents which stress reasons why a sustaining effects study is both desireable and practical. While the Chapter I legislative requirement requiring long term studies of compensatory programs will of course serve as a motivator, it is felt that districts will be more amenable to long term studies if they can be persuaded that the (1) benefits which study data can provide to local and state policy makers will offset study requirements, (2) studies can be completed with a minimal allocation of available district resources and finally, (3) State technical assistance and support for local districts is seen as being just a telephone call away.

Bibliography

1. U.S. Office of Education. Educational Programs That Work. 7th Edition, Fall 1980. Washington: U.S. Office of Education, 1980.

2. Connecticut Facilitator Project. Good Ideas in Education. New Haven: Area Cooperative Educational Services, 1981.

3. Connecticut State Department of Education. Handbook for Programs Funded Under Title I of the Elementary and Secondary Education Act. Preliminary Edition. Hartford: Bureau of Elementary and Secondary Education, 1978.

4. Model Programs Reading: Intensive Reading Instructional Teams. Publication OE-30034. Washington: U.S. Government Printing Office, 1971.

5. Huck, S. W., Cormier, W. H., & Bounds, Jr., W. G. Reading Statistics and Research. New York: Harper & Row, Publishers, 1974.

6. Tallmadge, G. K. The Joint Dissemination Review Panel Idea Book. Washington: U.S. Government Printing Office, 1977.