DOCUMENT RESUME

ED 058 296 TM 001 011

TITLE	A Follow-Up of: Summer 1968 Reading Institute Model School Division Washington, District of Columbia.
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	Nov 69
NOTE	25p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Educational Innovation; Elementary School Students; *Elementary School Teachers; *Followup Studies; Instructional Improvement; Interpersonal Competence; Kindergarten Children; Language Skills; *Program Evaluation; Reading Achievement; *Reading Clinics; *Reading Instruction; Reading Tests; Scores; Summer Programs; Testing Problems; Testing Programs; Test Results; Test Selection
IDENTIFIERS	Cooperative Primary Tests

ABSTRACT

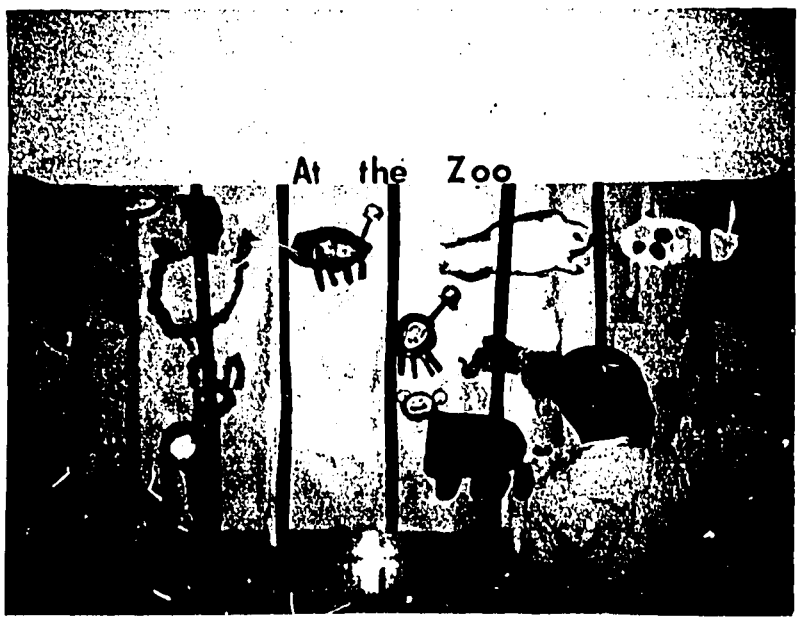
This report is the final one in a trilogy stemming from the 1968 Summer Reading Institute in the Washington Model School Division. The Institute was designed to meet the needs of elementary school teachers in their attempts to teach language skills and reading to students in grades K-3. This follow-up study examines the results of the Summer Reading Institute through an assessment of the growth of students whose teachers were participants in the Institute. (CK)

A Follow-up of:

Summer 1968 Reading Institute

Model School Division

Washington, District of Columbia



November 1969

Educational Testing Service

Princeton, New Jersey

TABLE OF CONTENTS

		Page
I.	General Background	5
II.	The Next Step	6
III.	Test Selection and Control Group Assignment	7
IV.	Workshops	9
V.	Test Administrations	10
	Table I - Number of Instances of Group and Individual Irregularities	11
VI.	The Test Data	11
	Table II - A Comparison of All Fall 1968 Cases with Fall 1968 Cases for Whom Spring 1969 Scores Are Available	12
	Table III - Precent Loss of Total N	13
	Table IV - Mean Gain in Scaled Scores	13
	Table V - Range of Mean Gains in Scaled Scores	14
	Graph I - Kindergarten Listening Pre-Post Tests	15
	Graph II - Grade 1 Listening Pre-Post Tests Grade 1 Reading Post Tests	16
	Graph III - Grade 2 Reading Pre-Post Tests	17
	Graph IV - Grade 3 Reading Pre-Post Tests	18
VII.	Discussion	19
VIII.	A Final Word	22
	APPENDIX A - Excerpt from Handbook - Cooperative Primary Tests	25

General Background

This report is the final one in a trilogy stemming from the Summer Reading Institute of 1968 in the Model School Division of the District of Columbia Public Schools, Washington, D.C. which involved a total of 52 Kindergarten, Grade 1, Grade 2, and Grade 3 teachers. For a complete understanding of the structure and objectives of this Institute and the follow-up activities the other two reports should also be consulted. 1/

In her history of the origin of the Institute, Miss Epstein noted, "The Institute was designed to meet the needs of the elementary teachers in the Model School Division. These teachers were responsible for teaching children to read and they, as no one else, knew they were failing." 2/ As the teachers had diagnosed their problem, they needed a three-point basic attack:

1. Additional budgets for the purchase of materials to be used in their language program for the 1968-69 academic year.

2. Exposure to a variety of approaches and techniques for the teaching of language skills and reading. The teacher, as the individual with expert knowledge of his younsters' learning styles, wanted to be able to select those approaches and techniques with the most promise for his youngsters.

3. The inter-personal skills and sensitivities of the teacher were to be heightened by systematic group experiences and training during the Institute. In the planning for the Institute it was decided to directly involve not only the classroom teachers themselves in the evaluation of the Institute, but also members of the Innovation Team, a group of dedicated teachers on special assignment serving fourteen inner-city schools as stimuli to improving educational practices.

In a very real sense then, there would be as many experimental treatments applied as teachers participating in the project.

1/ Michael Rosenfeld, An Evaluation of a Summer Reading Institute, 1968, PR-69-3, Educational Testing Service, March 1969.

Barbara Epstein, A Summer Institute in Teaching Beginning Reading, A Report of a Cooperative Project, The Model School Division District of Columbia Public Schools and Education Development Center, Raymond School, Washington, D.C., June 24 - August 2, 1968.

2/ Epstein, p.3.

No direct attempt was made to provide research designs for any part of the three-point basic attack which would result in conclusive statements concerning success or failure. A basic concern of the Innovation Team and the teachers was to answer first the practical question, "What happened?" To this end, the Rosenfeld report describes an attempt to "provide as simply as possible documentation of what the participants (teachers) felt they had learned during the Institute as well as their overall evaluation of it." In addition, the Epstein report provides a description, from the point of view of an outsider, of what actually took place in the Institute as a program. It was recognized that the question of how much effect the Institute training of teachers would have on the performance of students would be a far more difficult process than the purely descriptive ones noted above. Much of what was happening in the classroom was far too intangible and far too varied from classroom to classroom for a traditional research design to produce a dependable base of information for generalization. Nevertheless, by the conclusion of the Summer Institute in August 1968, everyone associated with the Institute felt that the program had a potential for such significant effect on teachers and students that some effort should be made to monitor the reading and language attainments of students in the classrooms of the Institute participants during the 1968-69 school year.

The Next Step

In August 1968, Miss Edith Baxter of the Innovation Team and Miss Barbara Epstein of the Pilot Communities staff, Education Development Center, contacted Educational Testing Service to obtain assistance in viewing the results of the Summer Reading Institute through an assessment of the growth of students whose teachers had been involved as Institute participants.

The request made to ETS incorporated the following assumptions:

1. Teachers regard most standardized tests as threatening to them personally and as inadequate tools for assessing the needs and the development of skills for their children.

2. In part at least, these circumstances exist because teachers have little experience with and inadequate information about testing--how to administer tests, how to assess the test results, and how to utilize those results.

In light of these assumptions, it was felt that in-service training on testing and on the utilization of its outcomes would be of value to the Institute-trained teachers and should be a facet of any attempt to use tests in their classrooms.

Test Selection and Control Group Assignment

A first task was to select the tests which would be used, those which would be most acceptable to the teachers, and which would serve an initial objective of providing information to the teachers in a form which would help them to assess the needs of their children and to feel more competent in applying some of the eclectic methods learned in the Institute. With these broad objectives in mind, tests were examined in the light of the following criteria:

1. An opportunity should be provided to examine as wide a range of language skills as possible recognizing as a fundamental tenet of the Institute that the teaching of reading and the development of language skills were two integrally related tasks.

2. Reading should be eliminated as a barrier to demonstrating ability in other areas (i.e., the need to learn something about the growth of children who had not yet learned to read).

3. An opportunity should be available to make maximum use of test data in providing feedback to the teachers who were responsible for improving the reading and language skills of the child.

4. Tests should have available statistically equated alternate forms for use in the test-retest situation.

5. Since the students' grade levels precluded any substantial experience with standardized tests, practice materials which would familiarize students with the mechanics of test-taking should be available.

While no test seemed ideally appropriate, the Cooperative Primary Listening Test and the Cooperative Primary Reading Test offered possibilities of matching the criteria and were subsequently adopted to be administered on the following schedule:

Grade	Fall 1968	Spring 1969
Kindergarten	Listening, Form 12A	Listening, Form 12B
Grade 1 [3/]	Listening, Form 12A	Listening, Form 12B Reading, Form 12B
Grade 2	Reading, Form 12A	Reading, Form 12B
Grade 3	Reading, Form 23A	Reading, Form 23B

The Cooperative Primary Tests provide a Pilot Test (see pp 9-10) and alternate forms which are equated through a scaled score system where the scaled score mean and standard deviation for the norms sample for the first semester of grade 3 were defined to be 150 and 10 respectively for each test.

The Listening Tests are based on a complex classification system which views listening as far more than merely receiving the spoken word, but as an act which includes comprehension, recall, and interpretation. A similar comprehensive classification system is also available for the Cooperative Primary Reading Tests (See Appendix A for a detailed statement of the "ground-rules" used in developing both the Listening and Reading Tests.)

To provide a baseline against which to interpret the test data for the students of Institute teachers, 23 non-Institute teachers were requested

---

[3/] Because most Grade 1 pupil's reading skills would substantially develop during Grade 1, but would be minimal or non-existent at the beginning of Grade 1, only listening skills were tested in fall 1968, but both listening and reading were tested at Grade 1 in spring 1969.

to also administer the reading and listening tests on the same schedule as the Institute participants and to participate in the in-service programs on testing. The proportion of Kindergarten, Grade 1, 2, and 3 teachers was the same as the proportion of Institute teachers in each of these four grades. The classes of Institute teachers were designated experimental groups; the other classes were designated as control groups.

Workshops

In September 1968, a four-man ETS team spent two days in the Model School Division offering in-service programs for both the teachers of experimental groups and the teachers of the control groups. In order to keep the size of the in-service training groups as small as possible, the in-service program was offered for two days, but any one teacher was involved for only a single day. During the September workshops, the ETS team first administered the Cooperative Primary Pilot Test to each teacher to familiarize teachers with the format of the tests and with the item types used. This was followed by a detailed presentation of good test administration procedures. Also covered in these workshops was the detailed discussion of the classification system for the listening and reading tests as presented in the Cooperative Primary Tests Handbook.

In November 1968, a second half-day workshop was held by one ETS staff member, at which the results of the fall 1968 testing were presented to the teachers. There were give-and-take discussions concerning the use of item data in diagnosing individual student difficulties and on the utilization of group data for the planning of instructional programs.

Test Administrations

Each youngster in the experimental and control groups was provided with the experience of taking the ten-item Cooperative Primary Pilot Test prior to sitting for the actual testing sessions. The Cooperative Primary Pilot Test is composed of extremely easy items, and it was felt that if a youngster could not respond successfully to at least three of the ten items on the Pilot Test, the actual testing session would be too frustrating an experience for him. It was further assumed that any youngster unable to cope with the Pilot Test would receive a zero score were he forced to participate in the actual testing session. Based on this assumption all youngsters unable to successfully complete the Pilot Test were assigned a zero raw score for the fall 1968 test administration. Approximately 17% of the Kindergarten, approximately 7% of Grade 1, and less than 1% of Grades 2 and 3 youngsters (representing approximately 4.4% of the total population) were so classified. In September 1968, the A forms of the tests were administered to all youngsters, experimental and control groups, who had successfully completed three items on the Pilot Test. The tests were then scored by ETS and the booklets and rosters of scaled scores were returned to the classroom teachers. In late April and early May 1968, the B forms of the tests were administered to all the youngsters in both experimental and control groups.

Provision was made to have teachers identify and report any irregularities which could invalidate test scores. Irregularities which concerned only an individual child (i.e., illness) were classified as individual irregularities. A circumstance which concerned the entire class (e.g., a general classroom disturbance) was classified as a group irregularity. If the irregularity was judged by the ETS Program Director to be sufficiently serious to invalidate

the test scores, it was labeled a Code 2 irregularity, so noted, and those test scores were excluded from all summary statistics. If the irregularity was judged not sufficient to invalidate the test scores, it was noted on the roster for the teacher, but the scores were included in the summary statistics. The most frequent Code 1 group irregularity was a teacher's questioning of the tests as too difficult for his students. The extent of group and individual Code 1 and 2 irregularities are summarized below.

TABLE I
NUMBER OF INSTANCES OF
GROUP and INDIVIDUAL IRREGULARITIES
REPORTED BY GRADE BY GROUP
(Note: A Code 2 Classification Invalidated Scores)

		ADMINISTRATION DATE							
		FALL 1968				SPRING 1969			
		Group Irregularities		Individual Irregularities		Group Irregularities		Individual Irregularities	
GRADE	GROUP	Code 1	Code 2	Code 1	Code 2	Code 1	Code 2	Code 1	Code 2
Kindergarten	Control	—	—	2	3	—	—	2	3
	Experimental	—	—	5	5	—	—	5	2
Grade 1	Control	—	—	8	—	25	—	4	—
	Experimental	67	—	9	—	65	—	3	4
Grade 2	Control	—	—	1	1	3	1	2	7
	Experimental	56	—	7	—	156	—	21	3
Grade 3	Control	—	—	—	—	—	—	—	3
	Experimental	—	—	5	4	1	—	10	7

The Test Data

During the fall 1968 processing of the tests, each student was assigned an identification number. These numbers were used to match data from the fall 1968 test administration with the data from the spring 1969 test administrations. The fall test results based on all children tested in September and on those children for whom both fall and spring data were available are presented for each group at each grade level as Table II.

TABLE II
A Comparison of All Fall 1968 Cases
with
Fall 1968 Cases for Whom Spring 1969
Scores Are Available
(Note: All data are presented in scaled score units.)

	Fall 1968 Testing (All Cases)		Spring 1969 Testing (Matched Cases)	
	Control	Experimental	Control	Experimental
Kindergarten				
range	103-151	103-132	103-151	103-132
N	152	106	119	76
Mean	118.6	116.5	119.9	116.9
σ	11.0	9.1	10.7	9.0
Grade 1				
range	103-163	103-146	103-155	103-146
N	143	328	54	183
Mean	125.7	126.1	129.8	127.3
σ	12.4	8.5	12.2	7.9
Grade 2				
range	119-150	119-163	120-150	119-163
N	195	527	159	380
Mean	131.9	133.2	132.0	133.9
σ	5.3	7.4	5.0	7.2
Grade 3				
range	122-162	119-169	122-162	119-169
N	122	339	103	251
Mean	138.7	139.7	139.2	140.8
σ	7.5	9.3	7.6	9.2

Attrition for any number of reasons such as student mobility is always a factor in educational research. Therefore, provision was made to identify students for whom both fall and spring test scores were available.

It is interesting to note in Table II that <u>without exception</u> the fall testing mean scaled scores are higher for the matched cases (those students for whom both spring and fall scores are available) than for the total group tested in fall 1968. This may suggest that the lower scoring children tend to be more mobile. It might be hypothesized that this could have the effect of depressing the amount of growth measured since the lower scoring youngster in the fall has, through an effective instructional program, the greater potential for increased growth.

The percentage loss of cases for each group at each grade is presented in Table III.

TABLE III
Percent Loss of Total N

	Control Total N	Matched Cases N	% Loss
Kindergarten	152	119	21.7
Grade 1	143	54	62.2
Grade 2	195	159	18.4
Grade 3	122	103	15.5
	Experimental Total N	**Matched Cases N**	**% Loss**
Kindergarten	106	76	28.3
Grade 1	328	183	44.2
Grade 2	527	380	24.2
Grade 3	339	251	25.9

The results of the fall-spring testing for all students who remained in their original group, experimental or control, and for whom both fall and spring test scores were available are presented as Graphs I-IV. To summarize these data, the mean scaled score gain demonstrated between the fall and spring test administration by group and by grade based on matched cases only is presented as Table IV.

TABLE IV
Mean Gain in Scaled Scores

	Control	Experimental
Kindergarten Listening	12.2	13.5
Grade 1 Listening	5.7	10.1
Grade 2 Reading	7.1	6.9
Grade 3 Reading	7.1	5.5

Somewhat concealed, however, by the mean gain statistics is the variability of the gains made by individual classes. The range of mean gains for each group for each grade is presented as Table V.

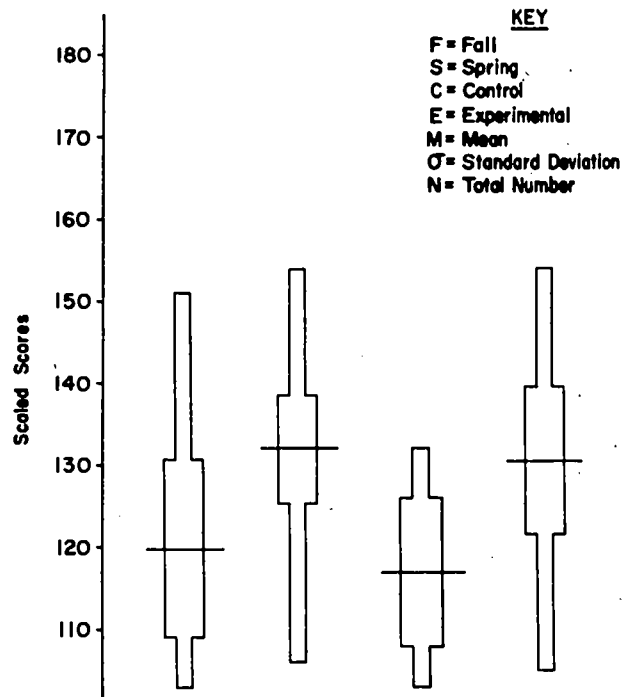
TABLE V
Range of Mean Gains in Scaled Scores

	Control		Experimental	
	Minimum Gain	Maximum Gain	Minimum Gain	Maximum Gain
Kindergarten Listening	6.9	29.1	3.3	23.3
Grade 1 Listening	-3.0	12.3	4.7	30.5
Grade 2 Reading	2.3	11.1	2.8	11.7
Grade 3 Reading	4.4	11.7	1.0	11.1

There appears to be no consistent, clearly indicated advantage for the students of the Institute teachers. However, the range of mean gains indicated for both control and experimental populations, assuming a reasonable comparability within as well as between groups, could lead to the conclusion often reached in instructional studies that the major variable contributing to the success or failure of an instructional program is not method, technique, or equipment, but, rather, the teacher.

As might be expected, listening achievement as measured by the Cooperative Primary Listening Test demonstrates an upward shift for both the control and the experimental groups at the Kindergarten level.

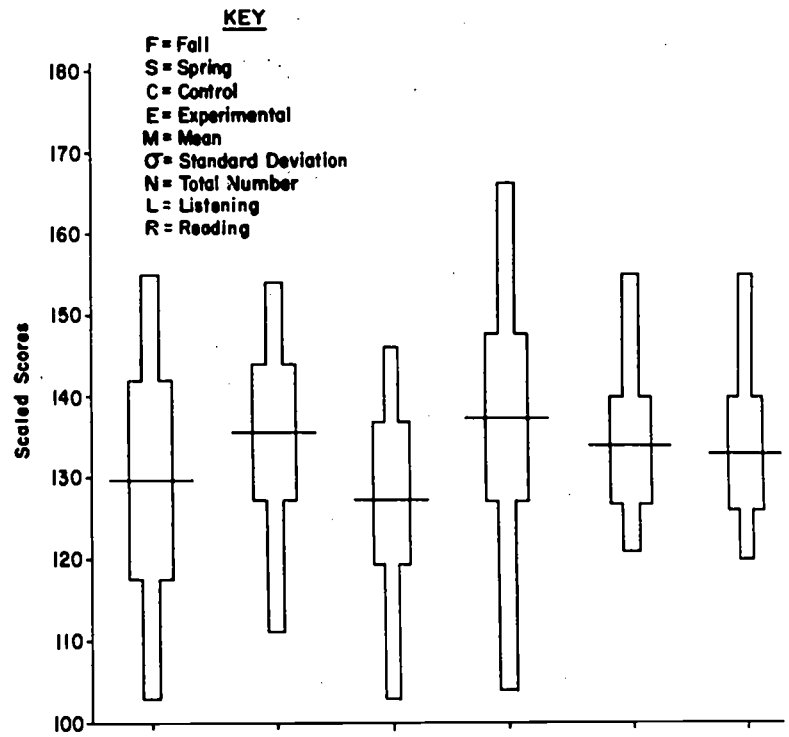
GRAPH I
Kindergarten Listening Pre-Post Tests
Matched Cases



Testing Group	FC	SC	FE	SE
range	103-151	106-154	103-132	105-154
M	119.9	132.1	116.9	130.4
σ	10.7	6.7	9.0	9.0
N	119	119	76	76

It must be noted, however, that the apparent growth may be exaggerated to an extent by the questionable appropriateness of the instrument itself for Kindergarten children. This observation is given added credence by the data for Kindergarten children for whom the Pilot Test served as the only test instrument administered during fall 1968. Graph I does, however, indicate that although there was an overall upward shift for both groups, the range of performance increased markedly more for the experimental than for the control group. If an objective of an instructional program is to enhance and increase individual differences, this may indicate an advantage for the Kindergarten students of Institute teachers.

GRAPH II
Grade 1 Listening Pre-Post Tests
Grade 1 Reading Post Tests
Matched Cases

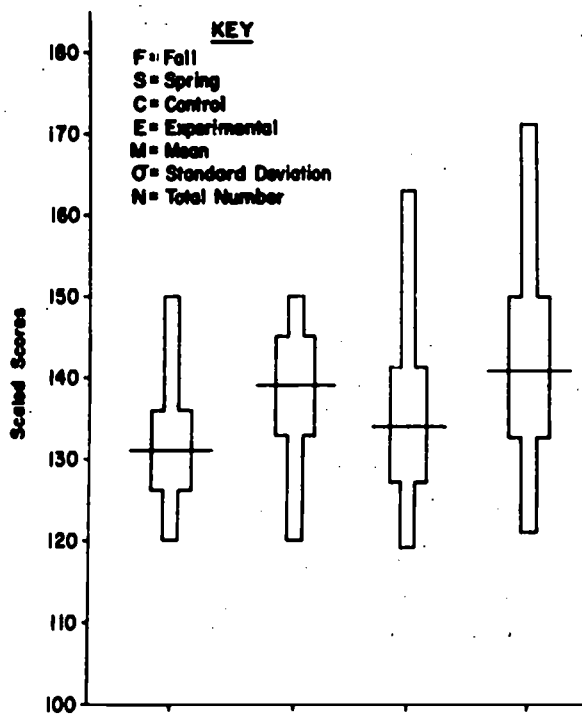


Testing Group	F CL	S CL	F EL	S EL	S CR	S ER
range	103-155	111-154	103-146	104-166	121-155	120-155
M	129.8	135.5	127.3	137.4	133.8	133.1
σ	12.2	8.4	7.9	10.3	7.3	7.0
N	54	54	183	183	54	183

As was observed in the Listening Test at the Kindergarten level, both the control and experimental group distributions for Grade 1 demonstrate an upward shift in listening achievement for the spring 1969 administration over the fall 1968 administration. Again, this is to be expected, but again, for the experimental group a greater range, and, for Grade 1 experimental students, a greater dispersion (standard deviation) is observable.

Reading as measured by the Cooperative Primary Reading Test administered in spring 1969 does not identify any significant differences in reading achievement between the control and the experimental groups at the end of Grade 1.

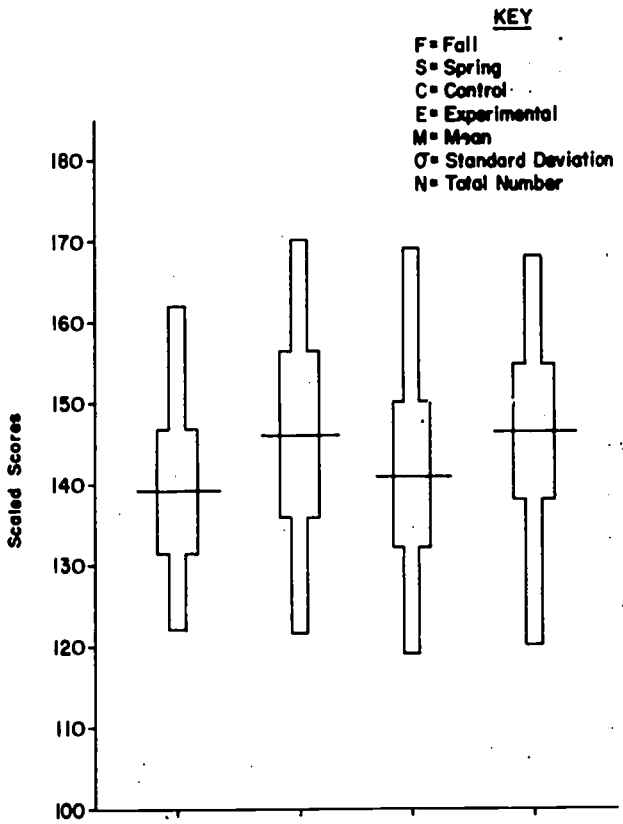
GRAPH III
Grade 2 Reading Pre-Post Tests
Matched Cases



Testing Group	FC	SC	FE	SE
range	120-150	120-150	119-163	121-171
M	132.0	139.1	133.9	140.8
σ	5.0	6.1	7.2	8.8
N	159	159	380	380

Reading achievement for Grade 2, as was true of listening achievement at the Kindergarten and Grade 1 levels, shows an overall upward shift for both the control and the experimental groups. In both groups the standard deviation increased for the spring testing indicating a wider dispersion of scores. The range of scores for the control group, however, remained unchanged from the fall to spring testing, while the lower as well as the upper limits of the total range increased for the experimental group at Grade 2. A slight difference, but nevertheless a continuance of the pattern observed at the two lower levels, is also found here, which may furnish an indication of a positive factor attributable to the Institute training the teachers of the experimental groups received during summer 1968.

GRAPH IV
Grade 3 Reading Pre-Post Tests
Matched Cases



Testing Group	FC	SC	FE	SE
range	122-162	121-170	119-169	120-168
M	139.2	146.3	140.8	146.3
σ	7.6	10.3	9.2	8.8
N	103	103	251	251

The fall and the spring test performances of both the experimental and the control groups at Grade 3 are remarkably similar (Graph IV). Reading achievement as measured by the Cooperative Primary Reading Test again demonstrates the expected upward shift in performance for both groups. The shift is not as dramatic as for the lower grade level groups, and the pattern of the increased range for the experimental group is not supported at the Grade 3 level. The hint that students of Institute teachers are benefited is not present at Grade 3. Indeed, the pattern, if it is one, is reversed with the control group's standard deviation increased and the experimental group's

standard deviation decreased for spring 1969 over fall 1968 test administrations.

Discussion

It has been clearly stated that the use of testing in the evaluation of the Institute was geared more to the needs of the teachers than to a demand for evaluation. Therefore, it is not to be expected that definitive statements regarding the impact of the Institute can be gleaned from these data. There do appear to be some trends, however, which are noteworthy. There are also some observations and questions which are probably significant in terms of planning for future research and evaluation and in planning for program development:

1. It appears from the data that the Institute may have improved the quality of instruction at the lower grade levels if it is accepted that an increase in individual differences is an indicator of more effective classrooms.

2. There is no indication that the students in the experimental classes were placed at any disadvantage by the more "open" classroom environment in which they were learning.

Such preliminary and limited conclusions are only tantalizing and point to the need for more thorough execution of research and evaluation and to the need for the development of more comprehensive and useful instruments for use with achievement test scores in assessing the usefulness of an instructional program. The data currently available would take on new significance, for example, if it were matched with systematic observations of all the classrooms, or if it related to a continuing study that matched subjects with programs in succeeding years, and if there were ways to attempt measures of growth in non-cognitive or skill areas which could have been fostered under different classroom climates. These data raise very interesting questions. If one infers that some effect was felt in the Kindergarten and

Grade 1 classes from the Institute, these questions remain:

> Why is this effect diminished in Grade 2 and not observable in Grade 3?
>
> Are the children locked into their self-concepts by the time they reach Grade 3?
>
> Are the teachers locked into images of controlled classes at this age level which makes it more difficult to utilize the free, open, and exploratory learning advocated by the Institute?
>
> Is it easier to change the image of the lower grade teacher toward openness than it is the higher grade teacher?

To answer some of these questions a battery of instruments and techniques to assess the role of teacher and student expectancy on creating a classroom environment is needed.

It would be overlooking a rich opportunity not to spell out some of the problems encountered in carrying out research and evaluation of schools, especially inner-city schools, which exercise constraints on design and possibly on the reliability of information. These environments particularly point to the necessity for vastly increased resources and skills to be applied in evaluation and research making it an active, ongoing integral, evolutionary part of a program. Some considerations which occurred during the course of this project are:

1. There is a need to recognize, as Educational Testing Service did clearly in the fall of 1968, that if a design is to be rigidly imposed for the sake of scientific validity, then the educator or the school may find itself with a program it no longer wants to know about or which does not meet the school's and the children's needs as originally defined. For example, if ETS had attempted to impose an arbitrary research design on an already ongoing program, it could have effectively sabotaged any good that the program might do. At times it would appear that there is a need in educational research to declare a legitimate "uncertainty principle." One simply cannot know at one and the same time what is going on and why it is going on, and still be able to make a quantitative judgment between two or more courses of action. Validity for one kind of information may logically rule out the validity of the other.

2. There sometimes may be a conflict between goals of research and the application of known findings. For example, the question of the control teachers versus the experimental teachers raises an issue. Would it not have been more scientifically desirable and provided more valuable information if, in the beginning of the program, teachers had been randomly assigned to control and experimental groups? Perhaps, but at the same time this approach would have conflicted with the stated wish of teachers to have a choice in the matter. It would also have precluded the possibility of the greater commitment to new programs possible when coercion is not involved in the decision to participate. A decision to have teachers volunteer certainly may be reflected in the findings in an indirect way and in a sense invalidate some of the information. For example, some older teachers may have selected classes in their individual schools. Older teachers tend not to volunteer for summer institutes. If this is so, then the comparative growth of the students of Institute teachers may have been masked by the circumstance that some control teachers were teaching children of generally higher ability. If this is so, of course, some of the growth shown in the classes of the experimental teachers is perhaps even more significant.

3. The uncertainty of conditions and the crisis orientation of inner-city schools makes systematic research, study, and evaluation extremely difficult. For example, an examination of Table III indicates the higher rate of turnover which occurs in many inner-city classrooms. The loss of students in the program in an effort to match scores does not, of course, reflect all personal moves of children. It also reflects administrative and organizational changes within the schools. The transitory nature of the child's position in these schools is a paramount concern and it makes matching and the valid comparison of results at best extremely difficult.

Table I through the tabulation of irregularities illustrates a further difficulty in obtaining data. Life in these schools is intensely volatile, changing, and stress-ridden. The teacher absentee rate is high. Programs may change radically as a result of these pressures. These facts simply point up the difficulty of attempting to impose systematic and highly structured research on the already overtaxed situations of inner-city schools. They also emphasize the need for caution in accepting any data obtained from these often "hectic" environments where "research" may come into conflict with school objectives or with the system, resulting in the lack of adequate

follow-through. To illustrate, long range planning was undertaken by the Innovation Team prior to the Institute, and discussions with principals and other administrators had resulted in assurances that control and experimental classes would not be included in the MSD's regular testing program and that Institute teachers would be permitted to make a whole hearted commitment to their programs. These assurances were not always honored. In one school, teachers had to engage in a high priority reading project which not only took precedence over the Institute program, but was in direct conflict with the approaches advocated by the Institute. In other instances, test scores may have been depressed by the administration of standardized achievement tests just prior to the administration of the Cooperative Primary Tests used in this project. What effect, if any, this actually had on the performance of the youngsters in the control and experimental groups is not determinable from this study. In those instances where teachers noted this circumstance existing, the scores for the youngsters involved were classified as Code 1 irregularities. There is, however, no indication as to the proportion of the classes for which the dual testing was a reality, but for which the teachers did not report any irregularities. It should be pointed out that some of the failure to honor assurances was not in any sense an overt attempt to sabotage the Institute's activities. It was simply one of the many examples of how multiple programs and multiple directions in our contemporary school bureaucracies often unknowingly conflict and unknowingly thwart the efforts of new programs.

A Final Word

Since what was being requested was not, strictly speaking, a research program, it was agreed that the ETS involvement would be through its General Programs Division rather than through one of its research divisions.

Innovation Team members and the EDC representatives agreed upon the following:

The teachers who had participated in the Institute would be regarded as an experimental group. It was recognized that the following factors would be likely to influence their profile as an experimental group:

a. They had volunteered for the Summer Institute.

b. From the initial number of volunteers a specific population had been chosen by the Innovation Team to provide adequate representation from each school and grade level.

c. A number of primary teachers, Kindergarten through Grade 3, equal in proportion to the proportion at each grade level in the experimental group was chosen as a control group. It was recognized that their profile as a group might be characterized by the fact that more experienced teachers tend not to enroll in summer institutes.

It should be pointed out that the phase of the evaluation which focused on student achievement through the utilization of tests was not regarded as a part of an overall design for evaluating the Institute-related activities of the previous summer. The plan to acquire and tabulate test information on the achievement of the children was regarded as an activity to provide corollary information for detecting the effects of:

1. Group training on teacher's behavior as reflected in measured student achievement.

2. The effects of additional equipment in language arts materials in the classrooms on measured student achievement.

For example, an answer to the question, "Does group training for teachers result in a more positive attitude toward reading?" could only be <u>inferred</u> if there was a significantly higher level of achievement demonstrated through test scores for the students in the experimental groups. However, even under these circumstances it was recognized that there would be no way to demonstrate that the control teachers were actually less, or more, skilled in group training than were the experimental group teachers. Therefore, it should be

made clear that the decision to look at children's test scores was made because of program requirements in the form of the requests of teachers to know how their children perform in relationship to other children and because of the previously stated opinions of teachers regarding testing and its significance. As evaluation the effort is considered significant in that it is a teacher-determined evaluation with more attention given to the needs in the growth of the teachers and the children than to the requirements of pure design for research and evaluation.

APPENDIX A

An Edited Excerpt from pages 7-9 of Handbook - Cooperative Primary Tests published by Cooperative Tests and Services, Educational Testing Service, Princeton, New Jersey, Copy right 1967.

TEST PERFORMANCE AND HOME BACKGROUND

The tests are predicated on the assumption that a principal aim of primary schooling for children from all backgrounds is to develop basic verbal and quantitative skills. However, it is recognized that the basic skills a child has developed by, say, the end of the first grade cannot be attributed wholly or even in the largest part to what the school has accomplished. The important thing, from both the teaching and measurement standpoints, is not how did a child come to be the way he is (and "how" is usually interpreted to include a large home background component) but where is he now--and what can be done by the school to strengthen his weaknesses and reinforce his strengths.

With this point of view, the answer to such a question as "How useful are these tests with 'disadvantaged' youngsters?" must be "As useful as teachers can make them, in terms of translating knowledge about pupils into appropriate learning activities." Some of the characteristics of the Cooperative Primary Tests designed to make the testing situation as fair and valid as possible would seem to have special relevance for children who come from homes where books, pictures, paper, and pencils are not standard items and where a standard brand of English is not spoken. These include elimination of reading as a barrier to showing abilities in some other areas, provision for adequate practice experience, and emphasis on measurement of improvable skills.

THE TESTS AND TEST QUESTIONS

Pilot Test. The 10-item Pilot test is designed to give children practice with the format and the kinds of questions and responses they will encounter in the regular tests in the series. It is recommended that the Pilot test be used prior to administration of any of the 12A and 12B forms (i.e., with pupils at the end of first grade or the beginning of second grade). In addition, teachers may want to use it with older children who have not experienced standardized tests before or who they feel may be likely to have trouble with the directions presented by the other Primary tests.

While experience with the Pilot test in pretesting and norming situations has indicated that almost all children can answer almost all items on the practice test, or at least understand what they are supposed to do, the teacher may occasionally find a child who does not seem to be able to handle the tasks it presents. If, after a second trial with the Pilot test at a later time, this still seems to be the case, the teacher is probably well advised not to go ahead to administer other tests in the series to this child. Interpretations from the other tests might be more misleading than helpful.

Listening. 'Listening,' as used in the title of these tests, means more than receiving the spoken word. It includes comprehension, recall, and interpretation.

The Listening tests are designed for presentation by the child's regular teacher. In other words, they are tests of face-to-face listening comprehension in the kind of situation the child must meet every day of his school life. The more standardized procedure of using tape recordings for presentation of the test was considered but rejected, because the recorded voices might speak in accents relatively strange in some sections of the country, classrooms vary in acoustical properties and students in some parts of the room might have difficulty hearing the recording, taped test material is more expensive and troublesome for all concerned, tapes are not adaptable to the natural interruptions and distractions of the classroom, and, most important, such presentation would test only ability to listen to recorded, disembodied voices, an interesting activity but one in which the pupil engages very infrequently during the normal school day.

Here are some of the ground rules that were adopted in developing and using the classification scheme for Listening:

1. Distinctions are made between concrete and abstract words on the basis of objects or entities the child can see, on one hand, and ideas, composites, actions, or descriptions, on the other. Thus, words labeled as concrete include web, magnet, and architect; words labeled abstract include balance, blizzard, abandon, and surrounded.

2. Any stimulus containing at least two sentences is labeled a paragraph.

3. Distinctions are made between comprehension of meaning in terms of illustration. Thus, a child might show comprehension of the word pierce by selecting a picture of a needle (as opposed to a picture of a hammer or a spoon) or comprehension of the word monument by selecting a picture of a monument (as opposed to a picture of a medal or of a street sign).

4. Distinctions between "recall" and "comprehension" are made on the basis of the complexity and/or length of the stimulus, although it is recognized that both recall and comprehension are involved to some extent in the items.

"Recall" is applied to responses to paragraphs with sets of items or relatively complicated paragraphs with single items, while "comprehension" is applied only to words, sentences, and short, simple paragraphs.

5. Category III is interpreted broadly to include situations in which certain information is clearly stated and the child simply has to identify a reshaping or translation of it and instances which are clearly inferential or evaluative.

6. Within all categories items range in difficulty. (Difficulty, of course, may be a function of the content of the material and the answer choices presented.)

It has been stated that results of tests in the Cooperative Primary series should be useful to the teacher in his instructional program. Maximum usefulness will come from study of responses of children in the class to each item. What kinds of words and sentences present the greatest problems in comprehension or interpretation? In what situations do the children have the most trouble remembering what was said? Are they willing and able to make inferences--to "add" something compatible to the story? It will be noted as the

test is studied that, to the extent possible, each distractor or incorrect choice was selected to tell the teacher something about the nature of the child's misconception or lapse in memory or comprehension.

Reading. A parallel structure was adopted for the Reading and Listening tests. The Reading tests differ from the Listening tests in that the child reads the words, sentences, and paragraphs rather than listens to them, the majority of the responses are words and sentences rather than pictures, "recall" on the Listening tests becomes "extraction" on the Reading tests where the child has the stimulus material in front of him, and the vocabulary level is appropriately below that of the Listening tests.

It will be noted that the child always reads the stimulus, but on the lower level forms, reading skill is indicated in approximately 40 per cent of the items by his choice of a picture response. On the upper level forms, this percentage is only about 15 per cent.

The vocabulary level of the Reading tests is geared to that of standard primary reading programs but is not tied to any particular instructional materials or published vocabulary lists.

The same kinds of considerations characterized the assignment of items to categories as those listed for the Listening tests.

As with the Listening tests and the other tests in the series, maximum benefits from administering the Reading tests will come from careful study of children's responses to each item on them. Clues picked up from study of children's reactions to the items may point to particular areas where special instruction is needed. For example, children in the national norms sample experienced considerable difficulty with items where one of the tasks was to identify whether the story did or did not provide a certain type of information. Instructional emphasis on this point would seem well worthwhile if we are going to produce a generation of readers who can distinguish between what they read in the lines and what they read between the lines.

An attempt was made to develop distracters which would help in identification of particular children's reading problems. For example, consider these Form 12 items in terms of the reasons why a child might select the incorrect choices: