DOCUMENT RESUME

ED 140 227 CS 003 397

TITLE Reading Center: 1972-73 Follow-Up Study. No. 72.
INSTITUTION Broward County School Board, Fort Lauderdale, Fla.
PUB DATE Dec 74
NOTE 15p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Elementary Education; *Program Effectiveness; Reading Achievement; *Reading Centers; *Reading Improvement; Reading Instruction; Reading Programs; *Reading Research

ABSTRACT

This study was designed to determine whether third-year pupils in a reading center project maintained their reading gains 16 months after leaving the program. Subjects consisted of two control groups and an experimental group--a random sample of third-year pupils enrolled at the reading center during the fall of the 1972-1973 school year. Criterion reading tests were administered as part of a country-wide testing program, completely independent of the reading center. Results of the study indicate that participants' gains tended to be maintained 16 months after they finished the program and that black participants may have benefited more than did the white participants, especially in the area of reading comprehension. (JM)

READING CENTER: 1972-73 FOLLOW-UP STUDY

THE SCHOOL BOARD OF BROWARD COUNTY, FLORIDA

Report No. 72
December, 1974

ABSTRACT

The present report is a follow-up on third year pupils served by the reading center project during the 1972-73 school year. The original study showed pupils who participated in the project tended to score higher on reading tests than would apparently have otherwise been the case. The present study sought to document whether or not these apparent gains would still hold up about sixteen months after participants left the program. Another less rigorous study conducted several years ago had indicated such gains had held up. The present study also supported this contention. There was, however, in this study a somewhat consistent but nonsignificant (in a statistical sense) tendency for gains to be maintained better among black than white pupils.

The predecessors of this study helped focus national attention upon the local reading center program. The American Institute of Research (AIR) used the documentation which has been accumulated regarding the effectiveness of this program as a basis for identifying it as an exemplary program. The findings of this report basically reinforce the conclusions of the previous studies.

Additional supporting evidence regarding the effectiveness of this program was provided by a recently completed evaluation of the local ESSA program. Methods and materials developed and used at the reading center were applied to over 2500 local first and second year pupils identified as needing help in reading on the basis of low test scores. The evaluative evidence indicated that the ESAA pupils so served, performed better in reading than would have otherwise been the case. Without the reading center, such results might not have been obtained.

The conclusions of this report were designed to promote constructive and critical thinking. Even though the evidence was strong that the reading center was benefiting those served, the criterion for selecting participants was changed during the 1973-74 school year. Beginning then, those "most in need" rather than those "ready to benefit" from these services were served by the center. Although the ESSA evaluation seemed to indicate that pupils "most in need" could benefit from methods and materials developed at the center, no "hard" evidence is presently available to document whether or not the program is equally effective with pupils selected on the basis of the new criteria. The type of "performance data" presently used to evaluate the program cannot provide such "hard" evidence. In any case this programs recognition by AIR should serve to alert decision-makers at the local, state and federal level as to just how rare and uncommon it is for a program such as this to be able to document its success with "hard" evidence. Although it is important to find out if the program is equally successful in serving pupils on a "most in need" basis, the program's effectiveness with pupils meeting its previous criteria for selection points out that compensatory funding should be allocated to insure that all local pupils who can benefit from reading center services, do in fact receive these services.

Finally there is a need for further reflection by decision-makers regarding the contributions which can be made by those familiar with tests and measurements and evaluative techniques to substantiating and implementing local projects. For example, an evaluative technique which is useful for one purpose may be quite unacceptable when extended into another area. Also tests are not infallible and, if used as selective devices, they must be supplemented by other selective criteria. Otherwise pupils will be selected for project participation who should not be and others will be left out who should be included.

READING CENTER: 1972-73 FOLLOW-UP STUDY

INTRODUCTION:

This study's purpose was to assess the long run benefits of participation in the reading center program. The reading center program features a clinical approach with an emphasis upon diagnosis and individualization.* The reading center has been continuously evaluated by this department since the 1967-68 school year.** Results of previous studies were essentially supportive of the center in that it appeared participants made significant gains in reading skills which they still maintained one year after being in the program. This report is organized so that the main body is directed to the lay reader. The more detailed and tedious material is appended.

PROCEDURES:

The 1972-73 school year presented local evaluators with a somewhat unique opportunity to implement a more rigorous research design than had been possible in previous years. Because more pupils were eligible for project participation than the program could house, it was possible to randomly select a control group from the list of all eligible pupils.

The "magic" of randomization is that it provides an effective control for all extraneous factors which could be associated with differences between pupils' test scores. The 1972-73 study*** therefore provided a highly defensible assessment of at least the short term benefits of the project. The present study was "built in" to the design of the 1972-73 study. In addition to randomly selecting a control group (control group #1) a second control group was selected from seven schools which were not served by Title I. The pupils in the second control group (control group #2), were selected in September 1972 on the basis of screening tests administered by reading center personnel.

Because entry into the project of pupils in control group #1 was delayed until summer rather than denied, it was thought that control group #2 would serve as the comparison group for the presently reported longitudinal study.

The results of the 1972-73 study indicated that a random sample of the third year pupils served by the project attained a higher level of reading proficiency than would have been likely had they not entered the program. This increased proficiency was maintained four months subsequent to pupils dismissal from the program. The present study was designed to see if participants still maintained their advantage about sixteen months after leaving the program. In brief, had the gains "washed out" over time?

As it happened, only a few members of control group #1 took advantage of their opportunity to participate in the program during the summer. Thus participants could be compared with both control groups. All testing in the 1972-73 study was carried out by reading center personnel. The present study was strengthened in that the criterion reading tests were administered completely independently of the reading center as part of a countywide testing program. The standardized tests used in the countywide testing were not the same as those used by reading center personnel.

*Persons interested in a detailed description of the program may contact Miss Louise Sears, Reading Center, The School Board of Broward County, Fort Lauderdale, Fla.
**See Research Reports #15, 29, 35, 36, 51 and 67.
***Report #68.

LIMITATIONS

Because of attrition, the findings could only apply to pupils still enrolled in the county. The reading center had for several years served the schools from which control group #1 was randomly drawn. Because of this, comparisons between control group #1 and participants were quite likely subject to a limitation referred to by researchers as "contamination." Were pupils in control group #1 really the same as they would have been had the reading center not existed? It is plausible that these pupils received indirect benefits due to the reading center before, during and after the time of the study.

On the other hand, comparisons between control group #2 and participants were subject to a limitation researchers call "confounding." Even though statistical procedures were used to "control for" pre-test differences, factors such as socio-economic differences between the pupil populations of Title I and non Title I schools could have influenced differences in post-test performance. This "confounding" prevented one from being certain that the only difference between the participants and control group #2 was entry into the reading center.

Such limitations are inherent in almost any practical study of ongoing educational programs. In the present study, the fact that the two control groups were available provided an opportunity to see if findings of both comparisons were congruent. The study was also strengthened in that a previous study had indicated long run as well as short run gains. The congruence of the findings of studies conducted at different times and using different approaches provides what is probably the strongest case for drawing justifiable conclusions from the application of research methods in educational contexts. In the case of the reading center, yet another source of supporting evidence became available at the time this report was written. Methods and materials used by reading center personnel were bodily incorporated into the local Emergency School Aid Act (ESAA) reading program implemented during the 1973-74 school year. Results of evaluations of the ESAA program were therefore also pertinent to building a body of congruent findings regarding at least the short term effectiveness of the reading center program. These findings will be discussed in the conclusions section.

RESULTS

Comparisons involving the first control group were clear-cut in the area of reading vocabulary. The experimental group scored significantly higher than control group #1 on the reading vocabulary test. Although findings on the reading comprehension test tended to be favorable to center participants, results were not statistically significant. Further, one could interpret the favorable trends in either of two ways. One interpretation would be that differences between white pupils tended to "wash out", but black pupils who participated tended to maintain their "edge" over the control group. This tendency approached statistical significance. The test of simple differences, collapsing across races, also approached statistical significance.

Comparisons involving the second control group on the reading vocabulary test revealed the presence of a highly complex relationship involving race, treatment and pre-test scores. However, apparently pupils of both races who scored low on the pre-tests and most black pupils benefited from participation in the program. On the reading comprehension test, simple overall differences favoring participants were statistically significant. However, again there was a tendency for program participation to benefit blacks more than whites which approached, but did not reach, statistical significance.

CONCLUSIONS

The findings of this study seemed to indicate that participants' gains tended to be maintained sixteen months after finishing the program. There was some, but by no means conclusive, evidence that black participants may have benefited more in the long run than did white participants.

Further congruent evidence regarding at least the short run impact of the reading center upon county pupils was provided by another study completed in the fall of 1974.* Methods and materials used in the clinical-remedial program at the reading center were used in a preventive-developmental program which served over twenty-five hundred first and second year pupils from thirty local schools who had initially scored very low on standardized reading readiness tests. It was found that pupils who were so served appeared to perform better on criterion reading tests than would have otherwise been the case. Thus, yet another source of data pointed in the same direction as the present study and previous studies.

Beginning with the 1973-74 school year the reading center altered its screening procedures. Prior to that time the center essentially served pupils with low achievement who had the potential to do better. Perhaps it would be accurate to say that "readiness" was the criterion for program entry. In 1973-74 the program began to serve pupils on the basis of "most in need." Pupils with the lowest scores on standardized tests were the ones eligible to enter the program with little regard for other factors. It might be mentioned that the ESAA project and the reading center both served pupils on the basis of "most in need." Both therefore had to develop "readiness" components for pupils who were not sufficiently advanced to be exposed to the methods and materials which have primarily constituted the program component which the research department has evaluated over the years. The results of the ESAA project evaluation were encouraging. However, it has not yet been possible for this department to document the impact of the reading center program upon its current enrollees.

The research evidence has provided a strong case for concluding that pupils benefited from the reading center program prior to the change in selection procedures. Furthermore, the ESAA evaluation provided a strong hint that methods and materials developed and used at the reading center could be effectively "exported" to in-school programs serving primary level pupils who were identified as being at the lowest levels in terms of measurable reading skills.

Plans are currently underway to implement the types of in-depth research studies of the current reading center project as have been done in the past. Whatever the results of future studies, however, the weight of the evidence suggests strongly that a significant segment of local pupils have benefited from the reading center program. The recent national attention focused upon the reading center as a result of the American Institute of Research (AIR)'s efforts to identify exemplary reading programs should suffice to indicate how uncommon it is for such a program to be able to rigorously document its effectiveness in terms of pupil outcomes. This report further reinforces the documentation. Therefore, even if subsequent evaluations indicate the program is not as effective when implemented on a serving those "most in need" basis, the results of the research to date should be put to positive use.

There may well be a need for local, state, and federal officials as well as advisory groups, to pause and reflect upon questions such as "Are some pupils who could be helped by this program now being excluded by the new criteria?"

*See 1973-74 Final Technical Progress Report, Emergency School Aid Act (ESAA), Basic Grant - OEG-4-73-1493.

"Are group administered standardized test results really so infallible that they can serve as a good basis for identifying pupils most in need of help?" "Is there anything to the concept of maturational readiness and, if so, what, where, and how should programs be developed and implemented for pupils who are slow to develop?"

Another area which also deserved careful attention is the current emphasis upon attaining performance objectives as an evaluative device. Performance objectives provide a very beneficial means of validating what a program can be expected to achieve. For example, test results might show that pupils will gain a certain number of grade equivalent months after program participation. Others considering implementing such a program with similar pupils can expect to attain similar results. Given accurate information based upon research or previous experience with the same population and instrument, it also makes sense to set realistic objectives and make their achievement an important evaluative criteron. Simple analogies with toothpaste ads may clarify the logic of the above. Use of brand A by a certain population may be associated with an average of X number of cavaties per person over a specified time. This validates for other members of that population what they might expect from using brand A. Of course, to use this data to evaluate brand A would require knowing the normal rate of cavaties per person in that population over the same time span. The danger of using performance objectives lies in making unwarranted evaluative assumptions, often at the stage of specifying objectives. Persons who are not at home in the area of tests and measurements cannot be expected to be familiar with the many traps and pitfalls associated with trying to accurately measure changes in test performance.* Because of this, the extension of the claims of essentially validitational studies into the evaluative domain can often be demonstrated to produce erroneous conclusions. For example, the gain of a norm group cannot be compared with the gain obtained by a certain subpopulation selected on the basis of receiving low initial test results. Unless a test is perfectly reliable, negative errors of measurement will occur a disproportionate number of times in the selected group. The group as a whole can be expected with mathematical certainly to score higher on a second administration of the test. In brief, some of the "gain" between pre and post tests obtained by a group so selected is due to errors of measurement rather than program benefits. Furthermore, if projects use scoring low on one fallible standardized test as the single criterion for selecting pupils "most in need"; many children will be selected who actually are not as much in need as many children who are not selected.

The above statements only reflect on the nature of tests and measurements. However, they do lend support to the need for systematic program evaluation to assist in making major educational decisions which will influence the welfare of pupils in need of compensatory services. For example, a strong case can be made in terms of the present evidence for assuring that pupils demonstrating low achievement with a potential to do better receive the kinds of services offered by the reading center, as our evaluative evidence shows that these students do benefit significantly from the services. An equally strong case can be made for demanding that similar hard evidence be secured whenever possible to substantiate all similar projects at the local, state and national level. A correlary of this is that a hard look needs to be taken at various means of evaluating such projects in order that decision makers can be informed and confident about how much credence to place in evaluative claims.

*See for example, Chester Harris, ed, Problems in Measuring Change, The University of Wisconsin Press, Madison, 1967.

APPENDIX A

PROCEDURES:

Two control group were selected in the 1972-73 study. One control group (control group #1) was randomly selected from the pool of third year pupils eligible for entrance to the program. Available facilities could not accommodate all eligible pupils. It was therefore made strictly a matter of chance as to which pupils would not receive program services during the regular school year. The program was open to control group #1 pupils that summer. However, a check of records revealed that only five pupils in control group #1 entered the reading center that summer or during the subsequent year. Control group #2 was selected by reading center personnel who employed regular project screening tests during September, 1972, in seven nonproject schools. Third year pupils in the nonproject schools who met the screening criteria for project participation composed control group #2. The experimental group consisted of a random sample of third year pupils enrolled at the reading center during the fall of the 1972-73 school year.

Form A of the Gates Primary Reading Test and the Otis short form IQ test constituted the initial screening criteria used in September. The September, 1972, reading scores on the Gates test were used as pre-tests in the present study. Post test reading scores were obtained from a county-wide testing program carried out in May, 1974. The Comprehensive Tests of Basic Skills (CTBS), expanded edition, level 2 was the test used in the countwide administration.

Student I.D. numbers were used to match countywide test results with data records used in the preceding reading center study. Table I illustrates the attrition which occurred during the course of the evaluative period. Five of the students in control group #1 were eliminated because they later entered the program.

TABLE I
ATTRITION

	Original Sample	First Study (May)	Follow-up Study (May)
	N	N	N
Control Group #1	60	58	41
Control Group #2	128	116	95
Exp. Group	119	107	82

LIMITATIONS:

Comparisons between control group #1 and the experimental group were soundest from a methodological viewpoint. Membership in these groups depended upon random assignment from a common population. The principal limitation upon such comparisons centers upon the problem of contamination. The reading center had served project schools for many years. It was therefore questionable as to whether or not the children in control group #1 had not already received indirect benefits from the reading center. They may well have not scored as high, for example, on their pre tests had the reading center program not existed in previous years.

The principal limitation of comparisons involving control Group #2 was confounding. Differences between this group and the experimental group could have been due to factors other than participation in the reading center program. For example, the population factors which determined which schools would be designated as Title I schools could obviously be an independent source of differences in test

results between control group #2 and the experimental group. Although the same statistical procedures were used in making the comparisons, the interpretation of the analyses differs. In the case of comparisons involving control group #2, it was posited that pre-test scores were probably somewhat associated with other confounding factors. The statistical tests simply attempted to isolate the contribution of the program which was not confounded with race and the pre-test scores. In the comparisons with the first control group, pre-test differences and differences in the proportion of black pupils in each group could be regarded as arising due to sampling errors, rather than confounding factors.

As has been noted, both procedures had inherent limitations. Concurrence of results from both comparisons would, of course, provide the most acceptable evidence regarding the presence or absence of a significant program effect.

The problem of attrition simply limited the generalizations of the findings to pupils like the ones who could be located for re-testing.

One limitation of the preceding study was eliminated in the present follow-up study. The criterion test in the present study was not administered by reading center personnel.

Finally, it must be noted that the reading center program has been continuously evaluated since the 1967-68 school year.* A previous, less rigorous, follow-up study had indicated that the program had seemed to produce long term benefits to pupils. So long as the results of this study remained consistent with the previous one, the cummulative evidence would appear to justify drawing fairly strong conclusions from the present study, despite the aforementioned limitations.

RESULTS

DESCRIPTIVE:

Means (averages) and standard deviations for the three groups involved in this study are summarized in Table II. Groups are further subdivided by race.

TABLE II
Means and Standard Deviations (Raw Scores)

Groups			Gates Pre-Tests				CTBS Post Tests			
			Read	Voc	Read	Comp	Read	Voc	Read	Comp
Race	Treatment	N	M	SD	M	SD	M	SD	M	SD
White	Exp	34	23.3	6.5	12.1	4.5	22.7	7.6	23.9	10.0
White	Cont #1	19	27.1	7.4	13.7	4.7	23.5	8.4	26.1	10.6
White	Cont #2	54	22.1	7.4	11.5	5.0	22.1	8.5	22.7	9.6
Black	Exp	48	21.0	6.3	12.9	4.8	22.1	6.9	24.1	8.8
Black	Cont #1	22	24.5	6.5	12.8	3.3	19.9	7.2	19.5	9.1
Black	Cont #2	41	16.8	7.5	10.0	4.9	16.8	8.4	16.2	8.6

Table II indicates that the surviving members of control group #1 tended to score higher on the pre-tests than did other groups. It would go beyond the available facts to attempt to explain this phenomena. However, some of it was due to sampling error, because the same tendency was present in the preceding study independently of the attrition factor.** It was also the case in the

*See Research Department Reports Nos. 15, 29, 35, 36, 51, 67.
**See Report #68, page 3.

preceding study that control group #2 fell below the other groups on the pre-tests. Part of the reason for these pre-test differences favoring groups from project schools may have been the "contaminating" effect of the reading center. Neither group may have done so well had their schools not been served by the reading center over the years. The particularly low performance of black pupils in control group #2, may be of some substantive significance. Were fewer black pupils in project schools who were identified as needing help in reading performing as low as similar pupils in nonproject schools?

Study of Table II can help in interpreting some of the statistical findings which follow. For example, a tendency can be discerned for black experimental pupils to have done relatively better on the post tests than did white experimental pupils, notwithstanding pre-test differences.

STATISTICAL ANALYSES:

PROCEDURES

A version of a MANOVA program developed by Eliot Cramer was used in the analyses which follow.* A more detailed summary of the statistical tables is appended. A two factor analysis of covariance design was used which followed the basic logic of model reduction inherent in the MANOVA program. The two factors were race and treatment. Pre-test reading vocabulary and comprehension scores were covaried. The MANOVA logic for model reduction is one which procedes from the complex to the simple.

Three tests of significance were of relevance to the present study.

1. Test of the equality of cell regression coefficients:
   Essentially this is a test of a complex interaction involving both factors and the two covariates. Significant results would show that the post test results depended simultaneously upon how well pupils did on the pre-tests as well as their race-treatment group membership.

2. Test of the race-treatment interaction:
   This is a test for an interaction between the two group membership factors. For example, blacks in the experimental group might have fared relatively better than whites in the experimental group.

3. Test for the treatment main effect:
   This is a test of the simple difference between all members of the two treatment group.

The MANOVA logic dictates that proceeding with tests of simpler models depends upon rejecting the preceding tests of more complex models. If, for example, the model which best fits the facts (or at least cannot be rejected at the .05 level of significance) involves a complex interaction, the analyses cease.

EXPERIMENTAL GROUP VS. CONTROL GROUP #1

Analyses proceeded past the test of equality of regression coefficients. The race-treatment interaction approached significance (P less than .089) at the conventional .05 level on the reading comprehension test. This was primarily because black experimental pupils tended to fare relatively better than white

*Cramer, Eliot and Sherin, Richard J. MANOVA, Multivatiate Analysis of Variance. A program distributed by Clyde Computing Service, Box 166, Coconut Grove Station, Miami, Florida.

experimental pupils. Acceptance of this hypothesis is primarily a judgmental rather than a statistical decision. A table of adjusted reading comprehension mean scores is provided below for the elucidation of these who feel that this model may best fit the facts. The mean scores are adjusted for pre-test differences (as though all four groups scored the same on the pre-tests).

TABLE III
Adjusted Means For Reading Comprehension Interaction
Control Group #1

Race	Treatment	N	Adjusted Mean (Read Comp)
White	Exp.	34	24.3
White	Cont. #1	19	24.3
Black	Exp.	48	24.7
Black	Cont #1	22	19.0

One can see from the above that in terms of this model only black experimental pupils benefited from the program.

The test of the simple treatment difference was significant (P less than .013) for the reading vocabulary test and approached significance for the comprehension test (P less than .067). A table of adjusted means is provided below.

TABLE IV
Adjusted Means for Tests of Simple Treatment Differences
Control Group #1

Group	N	Adjusted Mean (Read Voc)	Adjusted Mean (Read Comp)
Exp	82	23.1	24.7
Cont #1	41	20.1	21.5

One can see from the above that acceptance of these models leads to the conclusion that everyone, regardless of race or pre-test scores, tended to score higher if a member of the experimental rather than control group. This simple conclusion was clearly warranted by the sequence of statistical tests in the case of reading vocabularly scores. In the case of the comprehension test the only conclusion that was warranted, if one adheres strictly to the .05 convention of statistical significance, was that the groups were not different. This would imply that participation in the reading center program was not associated with significant differences in reading comprehension test results. As was noted, it is really not a statistical matter as to whether or not the models depicted in either of the tables, or the model involving no effect, was the model which best fitted the facts concerning reading comprehension scores.

EXPERIMENTAL GROUP VS. CONTROL GROUP #2

The hypothesis of equal regression coefficients could not be accepted in the case of the reading vocabulary test (P less than .042). This indicated that pupils' performance not only depended upon their race and treatment group, but also upon how they scored on the two pre-tests. Such complex relationships are difficult to clearly depict. Table V reflects a rather novel attempt to clarify the interaction. The first row in the table illustrates what the mean scores on the vocabularly test would have been had all groups scored the same average on the two pre-tests as did the white experimental group (i.e. 23.3 and 12.1). The second row shows the estimated means in terms of all groups receiving

the same average scores on the pre-tests as did the white control group (i.e., 22.1 and 11.5) and so on.

The "If" columns depict each of the four groups and report their pre-test scores. The "Then" columns show how each group would have fared under the circumstances of having the same average pre-test scores as each of the "If" groups. As can be noted from Table II, the bracketed diagonal entries are the average scores the groups actually received.

TABLE V
Estimated Vocabulary Scores based upon the Pre-Test Scores
Received by each of the Race-Treatment Groups
Control Group #2

Group	"IF" Pre Voc.	"IF" Pre Comp.	"THEN" Estimated Scores White Exp	White Cont	Black Exp	Black Cont
White Exp	23.3	12.1	(22.7)	23.2	22.9	21.6
White Cont	22.1	11.5	21.8	(22.1)	22.1	20.7
Black Exp	21.0	12.9	22.6	20.9	(22.1)	20.0
Black Cont	16.8	10.0	19.0	17.5	19.0	(16.8)

The table indicates that the estimates for the black experimental pupils were higher than for the black control pupils, given all four actual permutations of the average pre-test scores. The differences were most pronounced under the circumstances of low scores on both tests (the black control group situation). Extension of these results beyond the table would have shown control group blacks would have been estimated to do better than experimental group blacks only when their pre-test vocabulary scores were high and pre-test comprehension scores low. If any black children actually existed who scored quite a bit higher on their initial reading vocabulary test than on their reading comprehension test, the above results would lead one to conclude their progress in reading vocabularly would not have been helped by the reading center program. More generally, the benefits of the reading center in terms of reading vocabularly seemed most pronounced for black pupils with low scores on both pre-tests.

Happily for the sake of clarity much the same sort of relationship existed among the two white groups, but in a more pronounced way. In this case also white pupils with high pre-test vocabularly scores but low pre-test comprehension scores would apparently have been better off outside the center. Conversely white pupils with low pre-test vocabulary scores, but high comprehension scores would have apparently benefited greatly from the program.

Table V was constructed about actual permutations of average scores, because projecting to the above extremes would be meaningless since few or no pupils actually scored so differently on the two tests. The general conclusion from Table V would be that experimental group whites who scored low on the pre-tests appeared to do relatively better after treatment than was the case with experimental group whites who scored high on the pre-tests.

The reader must be warned that at this level of complexity, it may well be advisable to simply regard the above results as "noise". The improbable can and will happen simply as a matter of chance. If the present study were repeated several times, and the same complex interaction emerged each time, it would warrant serious consideration. Then one could use the results, for example, to

select pupils for project entrance who would benefit most from the program. Similarly, one would exclude from the project pupils who would be better off elsewhere. As matters stand, the same relationship did not exist in comparisons with control group #1 or comparisons involving the comprehension test. There is no way to know if the relationship would emerge again in a repetition of the study. It may be best to interpret the results on the reading vocabulary test in Table V as simply yielding favorable projections for experimental group pupils, especially blacks, under most realistic instances.

The statistical tests involving reading comprehension scores yielded slightly more clear-cut results. The interaction between race and treatment approached significance (P less than .077). Table VI is provided as a service to readers who may wish to further interpret this result. Since, a similar interaction approached significance in comparisons involving control group #1, more credence may be added to this possible interpretation.

TABLE VI
Adjusted Means for Reading Comprehension Interaction
Control Group #2

Race	Treatment	N	Adjusted Means (Read Comp.)*
White	Exp	34	22.7
White	Cont #2	54	22.3
Black	Exp	48	23.4
Black	Cont #2	41	18.6

One can see from the above that acceptance of the interaction model implies that black pupils seemed to have benefited more from the reading center in the long run than white pupils.

The simple test of main effect differences was significant (P less than .037). From a strictly statistical standpoint, therefore, this model warrants more credence than the interaction model. Table VII depicts adjusted means for this model.

TABLE VII
Adjusted Means For Simple Treatment Differences
Control Group #2

Groups	N	Adjusted Means (Read Comp.)
Exp.	82	23.2
Cont #2	95	20.6

On the above interpretation given equal pre-test scores, pupils of both races scored higher, if in the experimental group.

*The adjusted means for this table depended upon fitting an equation to different data than that used in constructing Table III. This is why the adjusted means for the experimental groups are not the same in the two tables. The main value of the tables lies in showing the size of contrasts, not in the exact values derived from the estimating procedures. The same applies to Tables IV and VII.

SUMMARY

The basic trends in the data were favorable to project pupils. Apparently project benefits were not "washed out," even after the passage of about a year and one-half. Results were not sufficiently clear cut to warrant concluding that pupils "across the board" benefited equally from the program. There was some indication that long-term results of participation might have been greater for blacks than whites, especially in the area of reading comprehension. The white-black interaction tendency was not as consistent in the area of reading vocabulary. One of the comparisons (control group #1) involving this area was favorable to reader center pupils regardless of race. However, differences in pre-test scores as well as race, which were found to be associated with reading center participation, complicated the results of the second comparisons. Whether such a complex relationship actually existed or was just due to "chance" conjunctions of data cannot be known. In any event, the overall trends in even the complicated case were mostly favorable to reading center participants.

The available evidence was basically sufficient to conclude that participation in the reading center program was probably associated with better long term test performance. The complexities which appeared in the data, especially those involving differential benefits for blacks, were not firmly enough established by this study to warrant decision-making concern.

APPENDIX B

STATISTICAL TESTS

TABLE I

Tests of Equality of Regression Coefficients:

TEST	DF	First Control vs Experimental F	P less than
Read Voc			
Read Voc.	6,111	1.18	.322
Read Comp.	6,111	.80	.573
		Second Control vs Experimental	
Read Voc.	6,165	2.24	.042*
Read Comp.	6,165	.90	.493

TABLE II

Tests of Treatment - Race Interaction:

TEST	DF	First Control vs. Experimental F	P less than
Read Voc.	1,117	.93	.337
Read Comp.	1,117	2.94	.089
		Second Control vs. Experimental	
Read Voc.	1,171	1.57	.212
Read Comp	1,117	3.17	.077

TABLE III

Tests of Treatment Main Effect:

TEST	DF	First Control vs. Experimental F	P less than
Read Voc.	1,117	6.40	.013
Read Comp.	1,117	3.41	.067
		Second Control vs. Experimental	
Read Voc.	1.171	.80	.373
Read Comp.	1,171	4.41	.037

*Since the interaction was significant at the .05 level, the remaining analyses of reading vocabulary scores were superfluous from the standpoint of the MANOVA logic. In brief, the hypothesis that the interaction model held could not be rejected.