THE PRETEST IN EVALUATIONS OF HEAD START AND RELATED PROGRAMS: A METHODOLOGICAL STUDY

A Thesis

Presented to

the Faculty of the Department of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Jeanne P. Deschner

August, 1968

ACKNOWLEDGMENTS

The author expresses thanks to Dr. Paul Dokecki for his counsel and assistance throughout this research project. Thanks are also due to officials in the Aldine Independent School District and the Porter-New Caney Independent School District for their cooperation and help in locating children for the sample groups. Special thanks are extended to Gloria Sammons, Charles Kendall, and Ed Richardson, who assisted with the administration of the tests.

J.P.D.

University of Houston

Houston, Texas

August, 1968

THE PRETEST IN EVALUATIONS OF HEAD START AND RELATED PROGRAMS: A METHODOLOGICAL STUDY

An Abstract of a Thesis

Presented to

the Faculty of the Department of Psychology

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

---

by

Jeanne P. Deschner

August, 1968

ABSTRACT

The issue studied was the timing of the pretest in educational program evaluations. An attempt was made to determine whether pretests should occur on the very first day of the program, or whether they may be delayed, as has been done with some reported Head Start evaluations. The Slosson Intelligence Test was given to a random sample of children enrolling in Head Start on the opening day, followed by a posttest 3 wks. later. Control children were tested at comparable times. The E group as a whole gained slightly, but not significantly, more than the C group. However, the E boys gained significantly more than the E girls, whereas C boys gained somewhat less than C girls. The results suggested that the opening day of a program is to be preferred over later pretest dates, and that a pretest delayed beyond the second week would find subjects already modified by their contact with the program. Practice effect between pretest and posttest did not appear to be a problem, but some coaching of C subjects by their parents occurred between tests, since parents were allowed to witness the C pretests. Several possible causes for the surprising differences in IQ gains between E boys and girls were discussed in conclusion.

TABLE OF CONTENTS

CHAPTER PAGE

LIST OF TABLES

TABLE PAGE

TABLE PAGE

LIST OF FIGURES

CHAPTER I

INTRODUCTION

This study investigated the effects that the timing of the pretest might have on the evaluation of preschool educational programs such as Head Start.

Various forms of Head Start, the name given to the federally financed, locally administered preschool education programs for culturally disadvantaged children, have flourished throughout the United States for several years, as a kind of massive social experiment. Despite the evident need for research concerning the effectiveness of Head Start, the output of evaluative studies is surprisingly small. Moreover, when the sparse literature is reviewed, serious methodological problems become apparent. Some of the most vexing questions are those concerning pretest administration - when, where, and what to use for a pretest. When, the question of proper pretest timing, was chosen as the focus of the present study.

To be a genuine pretest, the subjects should be measured before the educational treatment ever starts. Certain pilot projects in preschool enrichment for disadvantaged children have been able to make random selection of experimental (E)

and control (C) subjects and pretest them all prior to treatment (e.g. Gray & Klaus, 1963; Weichart, 1967). Beller and Nash (1965) undertook to pretest all subjects prior to the opening of a standard summer Head Start program. The pretesting was a difficult task, because the children were very shy and the mothers were suspicious and confused by the testing process. The researchers felt that these conditions could have affected the reliability of the data obtained.

Pretreatment testing is probably out of the question for most actual Head Start programs for the following reasons: it would take great deal of manpower which is usually unavailable at that moment, home conditions are often unsuitable for testing, it is hard to know who the pupils will be before they actually arrive on the first day, and finally, the whole pretesting effort could easily arouse so much suspicion and misunderstanding that it might scuttle the very program it was designed to evaluate. In the face of such formidible difficulties it is usually preferable to wait until Head Start opens its doors and then give the pretest under the controlled conditions of the preschool setting.

The first day of the program would seem the logical choice for the pretest, since the subjects' contact with the program would be minimal. Strangely enough, not a single

study has reported an opening day pretest. This may be because it would be necessary to bring in special examiners, since the regular staff would all be fully occupied on the first day. Or it may be that none of the researchers have recognized the importance of proper timing of the pretest for the meaningfulness of later evaluations of the program, and so just failed to mention when they gave the pretest. But the timing of the pretest is a crucial matter because every day that it is delayed finds the subjects further modified by their experiences in Head Start, until some unknown date when the pretest can no longer be considered pre, and a major part of the program evaluation has lost validity.

Berzonsky (1967) waited one week before giving the pretest, on the theory that capabilities would be measured more accurately after the children had settled down in school. The problem, Berzonsky discovered, was to know when "settling down" left off and the effects of the Head Start treatment began. It turned out that the E group scored much higher on the pretest than the C group, suggesting that the one week had been enough to change the E subjects. He also noted a positive correlation between posttest IQ and the number of days the subject had attended Head Start.

After deciding when to pretest the E subjects, the researcher must settle another difficult question: when to pretest the C subjects. Selection and testing of a C group present such problems for most Head Start evaluations that the temptation is to do without a C group at all; however, at least two studies show that a comparison group is essential. Alpern (1966) and Tannenbaum (1966) in separate studies, pretested and carefully matched E and C groups, and discovered on the posttest that the C group gained as much as the E group which had received the special education. Furthermore, many studies (e.g. Wolff & Stein, 1965) have found that the relative advantage enjoyed by Head Start graduates diminshes almost to the vanishing point a few months after enrollment in the regular school. Analysis of the causes of this discouraging trend requires carefully matched E and C groups.

One solution to the need for a C group, which has been used in a few studies, is to make a post hoc C group out of the children who failed to attend Head Start and who later share classrooms with those who did. The Houston Independent School District (personal communication) asked first grade teachers to compare the performance of Head Start and non-Head Start, and kindergarten and non-kindergarten, children in their classrooms. Giles and Daniel (1967) in an otherwise

sophisticated enquiry into the influence of Head Start on children's language patterns, also tested first graders and divided them into E and C groups, after the treatment was all over. Eisenberg and Connors (1966) tried pretesting the C subjects in kindergarten, three months after the E subjects were pretested during a summer Head Start program.

The chief trouble with convening a post hoc C group from the E subjects' later classmates is that the researcher must assume without evidence that E and C groups were alike before treatment began, which is probably more likely to be untrue than true. Head Start programs are preceded by an intensive recruitment effort in the impoverished community, and in many localities this recruitment has been so successful that the only children who stay away are those who are ineligible because of higher economic status, are to sick or retarded to come, or else are from families so neglectful of the child's welfare or else so suspicious of the school that the child is not sent. Any of the above reasons would make Head Start stay-aways different from participants during later educational experiences as well, so that a post hoc evaluation might easily give credit to Head Start for differences really caused by subject selection. The only way to control for the possibility of subject selection differences is to give the same

pretest to both E and C groups.

The when and the where of the C pretest become more difficult questions when the decision is made to give the E pretest at school, for C subjects, by definition, are unaffiliated with any school at the time of the pretest. However, the timing can be held constant by pretesting both groups at the same time, as nearly as possible.

The present study explored the feasibility of conducting the E pretest on the presumed optimum date, the very first day of Head Start, and giving the C pretests the following week. A special team of examiners, not part of the Head Start staff, administered the individual pretests to subjects randomly selected from the Head Start enrollees on the first morning. The same team of examiners gave the C pretests the following week. Two to three weeks later the posttests were given the E group, and the following week posttests were given the C group, thus preserving the same time intervals.

The chief question to be answered by the present study was, would the E children show significantly greater gains than the C children in the short interval of two or three weeks of Head Start experience? If they did, it would establish a cut-off date beyond which a test could not be considered a pretest. On the other hand, if the E and C

did not show a significantly different rate of gain between tests, it would indicate that tests administered as late as the second or perhaps even the third week of Head Start might serve as E group pretests in later evaluations of the effect of the treatment. Such a negative result would be an important finding because it would ease the practical problems of timing pretests in Head Start and similar studies.

Thus, the study was conceived to explore one specific methodological question, the timing of the pretest. It was not the intent of the study to look into the question of what should be administered as a pretest, though this, too, is a problem which needs attention. Furthermore, it was not the purpose of this study to say anything about the effectiveness of Head Start, either for short-term gains in test scores or for long-term benefits; but was designed purely to make a contribution of the method of executing such evaluative studies.

CHAPTER II

METHOD

Subjects

Subjects for the E group were selected from children enrolled in the summer Head Start classes held at four elementary schools in the Aldine Independent School District, Aldine, Texas. Two of the schools had an all-Negro enrollment; the other two had a mixture of Anglo and Latin surnamed white pupils, plus a few Negro children. As a condition of enrollment, parents of each pupil had signed a statement attesting that the annual family income did not exceed the permissable ceiling for Head Start, which went as high as $5000 for families of eight or more. Aldine does not have a free kindergarten, so the Head Start classes were limited to children eligible for first grade in September but who had not attended kindergarten. Therefore the program was serving as the children's introduction to formal education, even though most of them were already six years old. The mean age of the subjects in the E group was 6 yr. 3 mo.

As they arrived on opening day the pupils were randomly assigned to classrooms by the Head Start staff. An examiner stationed at each of the four schools went to every classroom

in turn, picking a boy and a girl to be subjects by an alphabetical sort method. In all, 48 subjects were tested, about ten per cent of the total enrollment. This group shrank to 33, since 7 children were absent during the week of the post-test, and another 8 were subtracted after the examiner learned that they had attended a day care - kindergarten the previous year. These 8 subjects were considered separately as a "prior schooling" group. The 33 remaining subjects included representatives of all three ethnic groups attending the Head Start, and children of each sex were seen by examiners of each sex, so that an analysis of the effects of ethnicity and sex of subject upon test results could be made.

It was not possible to obtain C subjects from the same school district as the E group. Aldine officials felt that the Head Start recruitment had reached nearly all eligible children, and also did not wish anyone to go out into the community to locate any left behind, on the grounds that it might stir up trouble. This was not an unrealistic fear in view of the many disturbances among the disadvantaged groups all over America in the"long, hot summer" of 1967. The author contacted a number of adjoining and comparable school districts which, like Aldine, lacked a kindergarten, before finding one that had enough economically deprived children

along with a relatively peaceful atmosphere, to provide a locale for home testing of C subjects.

All C subjects came from the Porter - New Caney Independent School District, which serves two semi-rural communities, 10 and 15 miles north of Aldine. Though the communities are smaller than Aldine and somewhat more rural in atmosphere, local residents consider them similar to Aldine. Families in all three communities are mainly from the laboring class, with the breadwinner usually commuting into Houston for work.

It was not possible to conduct a formal investigation into the economic or social background of either E or C subjects because of the volatile general atmosphere already mentioned. No home visits were attempted by the Head Start staff to check on the accuracy of the income statements of pupils' families. Though the examiners did make home visits to each C subject, no attempt was made at data collection other than the pretest itself, again for fear of precipitating an incident which would damage the excellent working relationship with the people in the community and also the school administration.

All C subjects were tested at their own homes. A total of 28 children were pretested, but 5 could not be found the

week of the posttest. The final C group was composed of 23 subjects, 9 boys and 14 girls. This was a somewhat smaller group than had been hoped for, especially in view of the fact that three examiners devoted full time for a week to locating and testing the C subjects.

Head Start children were picked up outside their homes each morning by a school bus, and were delivered each afternoon at 1:30 to their own doorsteps. As soon as they arrived at school they received breakfast in the cafeteria, where they were seated by classes with their own teachers, teacher's aides (mothers) and volunteer aides (high school boys and girls). After breakfast all the children were taught to brush their teeth and wash their hands. A lunch was also served to the children each day, followed again by the cleanliness routines. The classroom curriculum was based on the kindergarten curriculum for the Aldine schools. Puzzles, games, singing, coloring and pasting, simple crafts, dancing, dress-up, and listening to stories and records were all part of the classroom activities. Each class also had thirty minutes of physical education each day, followed by a juice snack.

Some but not a great deal of training was specifically aimed at use of language and numbers, learnings which might be expected to increase intelligence test scores. Emphasis was placed on informal conversation rather than formal word

drill in the development of the children's language skills.

The daily routine had to be adjustable, for there were interruptions nearly every day. Each child received thorough medical and dental examinations, as well as individual and group psychological tests. All pupils went on field trips every two weeks; in the alternate weeks films and guest speakers were featured. It is possible to estimate about 5 hours of formal class instruction for each child per week, plus about 10 other hours of socializing activities such as eating, physical education, field trips, and bus rides, for a total of 25 hours of Head Start treatment per week per subject.

At the same time that the E children were attending Head Start, the C children were receiving an educational treatment of their own, associated with their own homes. It was assumed that this influence would be very similar to the home environment from which the E subjects had come. The June and July weather was very hot, and almost none of the C homes had air conditioning, so that the children spent most of their time out of doors, away from the influence of TV, and interacting with peers in the immediate vicinity of their homes. They appeared to have few toys but many animals to play with, and no organized community recreation.

A number of the mothers proudly showed sets of World Book and Child Craft to the examiners, relating how the teacher who had sold them the encyclopedias had also instructed them in the use of the various programmed instructional devices, and had urged the mothers to use the programs and to read to their children, in order to make up for the lack of a kindergarten in the community. No attempt was made to find out how much the mothers were actually using the books and aids, though evaluation of their use would have made an interesting study in itself.

Most of the C families had a garden and animals such as rabbits and dogs, but none of them were on working farms. Some of the houses were in varying stages of do-it-yourself construction. The over-all impression given by the C families was of an upper-lower class group with strong upward mobility, in many instances.

Procedure

Each subject randomly picked for the E group was taken by the examiner to an empty school room and given the pretest. Administration of the test took 10 or 15 min., then the pupil was accompanied back to his room and another subject chosen. Posttests consisted of the identical test, given in the same location to the same subjects, but most of the time by a differ-

ent examiner. One third of the E posttests were given in the second week of Head Start. The other 22 posttests were given administered in the third week. Means of the second week and third week tests were so nearly identical that they have been handled as a single posttest.

The C subjects were tested the week following the E group pretest, but with the planned difference that they were tested wherever a place with two chairs could be found in their own homes, which might be kitchen, living room or porch. A second difference arose in the course of pretesting the C subjects. Though the E children were all tested by themselves, the C children usually had to be tested in full view of mother and siblings and even occasional visitors. These "kibitzers" were asked not to prompt the subject and restrained themselves well, on the whole. They were not asked to leave for fear of arousing misunderstandings about what the examiner, who after all was a total stranger to the mothers, was going to do to the child. The C group was posttested three weeks after the pretest, which was the week following conclusion of the E posttests.

Examiners

The E subjects were pretested by two female examiners, both experienced in administration of other intelligence tests, and two male examiners who were inexperienced in test adminis-

tration. It was felt, however, that all examiners quickly mastered administration of the test used. Posttests for the E subjects were given 1) during the second week of Head Start by a team of five college students who were testing a large number of subjects as part of another study, and who included eleven E subjects in their sample, or 2) during the third week by one of the female examiners who gave the pretest. All examiners were Anglos. Three of the E group examiners, two male and one female, were available to conduct the C group tests. The same examiners saw the same subjects for both pre and posttests.

The Slosson Intelligence Test

The Slosson Intelligence Test for Children and Adults (Slosson, 1963), commonly referred to as the SIT, was used for both pre and posttests of all subjects included in the present study. The age range of the SIT extends from two weeks to 27 years old. Items for the very young children are taken from the Gesell Developmental Schedules; most of the rest are adapted from the Stanford-Binet. The chief difference is the brevity of the total test. The examiners found it took an average of 15 min. to administer and score the SIT, making it possible to see ten children in one Head Start day.

SIT IQ scores are quotients, derived by dividing the

subjects' total score, which is expressed as a mental age, by his chronological age. Tables in the SIT manual show consistently high correlations (.90 to .98) between IQs obtained from the SIT and from the Standford-Binet, Form L-M, and also with IQs derived from the Weschler Intelligence Scale for Children, for all ages.(Slosson, 1963,pp. v, viii) The one weakness of the SIT was the lack of new item validation statistics; the test author relied on the old Terman and Merrill standardization of the items, since most of the SIT items come straight from the Stanford-Binet without revision. However, in a few places items appear at a new point in the age scale, and all vocabulary items are different and are cast in a new form. It was felt that the SIT would be even more useful for Head Start research if a new item standardization had been done, using a generous sampling of culturally deprived subjects from various ethnic groups in the over-all sample.

CHAPTER III

RESULTS

A Lindquist Type I analysis of variance (Lindquist, 1953) was used with the total SIT IQ scores, to assess impact of the two or three weeks of Head Start upon the IQs of the subjects. The interaction between groups x pre-post testing was found to be nonsignificant, as shown in Table 1. Comparisons of the group means by *t* test revealed that both groups had made statistically significant gains ($p < .01$), with the E subjects gaining 6.21 IQ points and the C subjects gaining 4.95 IQ points. In addition, a *t* test for uncorrelated means (Guilford, 1965) showed that the differences between groups were statistically significant ($p < .01$), which would indicate that E and C groups had not come from strictly comparable populations. (See Table 2.)

Comparison of sex x group x testing through a mixed design analysis of variance was not done because of the unequal Ns of the groups. Sex effects within E and C groups were compared in separate analyses of variance, as shown in Tables 3 and 4. The interaction of sex x testing was statistically significant ($p < .01$) for the E group (see Table 3) but nonsignificant for the C group (see Table 4). These interactions are represented graphically in Figure 1, showing the

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE OF PRETEST AND POSTTEST SIT IQ SCORES FOR E AND C GROUPS

Source	df	MS	F
Between Subjects			
E - C Groups (B)	1	2068.68	7.51 **
Error (b)	54	275.49	
Within Subjects			
Pre - Post (A)	1	908.58	26.75 **
A X B	1	10.68	<1
Error (w)	54	33.97	

** $p < .01$

TABLE 2

MEANS, STANDARD DEVIATIONS, AND COMPARISON OF DIFFERENCES FOR PRETEST AND POSTTEST SIT IQ SCORES FOR E AND C GROUPS

Group	Test Mean				Gain	t
	Pre	S.D.	Post	S.D.		
E (N - 33)	81.63	11.1	87.84	12.3	6.21	3.91 **
C (N - 23)	91.00	16.9	95.95	12.4	4.95	3.11 **
Difference	9.37		8.11		1.26	
t	11.86 **		10.26 **			

** $p < .01$

TABLE 3

SUMMARY OF ANALYSIS OF VARIANCE OF PRETEST AND POSTTEST SIT IQ SCORES FOR BOYS AND GIRLS IN E GROUP

Source	df	MS	F
Between Subjects			
Boys - Girls (B)	1	268.40	1.11
Error (b)	31	242.18	
Within Subjects			
Pre - Post (A)	1	636.74	18.73 **
A X B	1	243.95	7.18 **
Error (w)	31	33.99	

** $p < .01$

TABLE 4

SUMMARY OF ANALYSIS OF VARIANCE OF PRETEST AND POSTTEST SIT IQ SCORES FOR BOYS AND GIRLS IN C GROUP

Source	df	MS	F
Between Subjects			
Boys - Girls (B)	1	357.76	1.11
Error (b)	21	321.08	
Within Subjects			
Pre - Post (A)	1	282.52	11.14 **
A X B	1	3.99	<1
Error (w)	21	25.36	

** $p < .01$

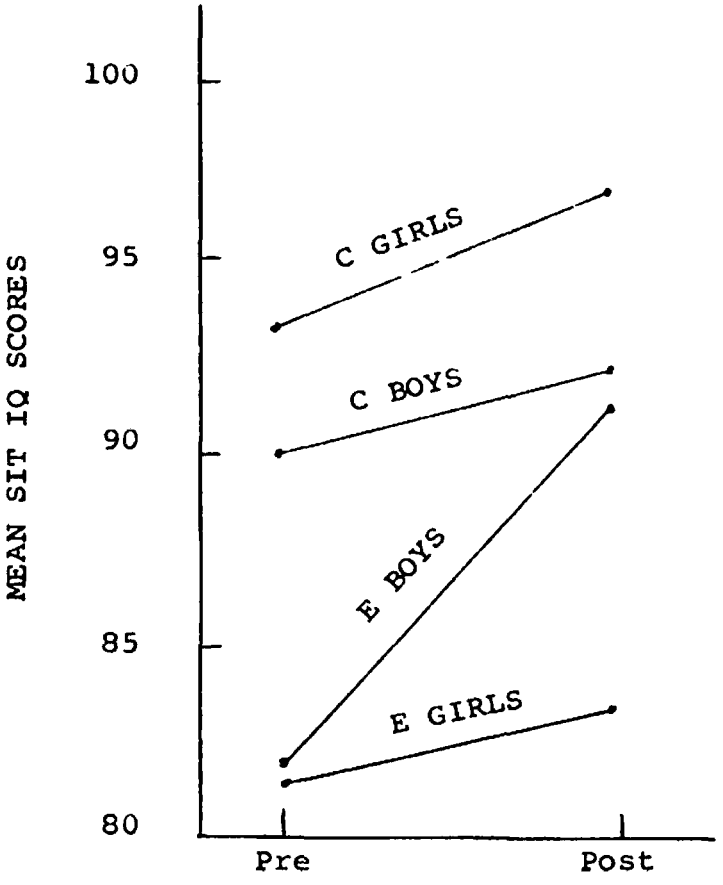

FIGURE 1

MEAN SIT IQ SCORES ON PRETEST AND POSTTEST

FOR E AND C GROUPS, AS RELATED TO SEX OF SUBJECT

steeper gains of the E boys, compared with the other sex groups.

Table 5 shows that *t* tests of mean differences indicated that E boys, C boys and C girls all made significant gains ($p < .01$), while the E girls, who as a group gained only 2.00 IQ points in two or three weeks of Head Start, did not make a significant gain.

Comparison of the means of the three ethnic groups represented in the E sample, shown in Table 6, revealed three distinct subgroups within the E group, each of which began at a different IQ mean, had a different variance, but made close to the same amount of gain. However, gains for the Latin subgroup did not reach statistical significance.

Since only four non-Anglo subjects were found for inclusion in the C group, an ethnic analysis of the C group was not made. The difference in racial composition of E and C samples probably accounts for a good deal of initial differences between the two groups.

Comparison of ethnic x sex groups within the E sample revealed still more differences, shown in Table 7. Among the Latin surnamed children alone, the girls exceeded the boys on the pretest, and gained almost as much as the boys on the posttest. Within Anglo and Negro groups, boys made large

TABLE 5

MEANS, STANDARD DEVIATIONS, AND COMPARISON OF DIFFERENCES FOR PRETEST AND POSTTEST SIT IQ SCORES FOR BOYS AND GIRLS IN E AND C GROUPS

Group	Test Mean				Gain	t
	Pre	S.D.	Post	S.D.		
E Boys (N - 18)	81.72	14.4	91.44	12.1	9.72	6.79 **
E Girls (N - 15)	81.53	5.3	83.53	3.9	2.00	1.39
C Boys (N - 9)	87.89	18.1	92.11	14.6	4.23	2.86 **
C Girls (N - 14)	93.29	10.7	97.00	11.2	5.42	3.66 **

** $p < .01$

TABLE 6

MEANS, STANDARD DEVIATIONS, AND COMPARISON OF DIFFERENCES FOR PRETEST AND POSTTEST SIT IQ SCORES FOR DIFFERENT ETHNIC GROUPS WITHIN THE E GROUP

Group	Test Mean				Gain	*t*
	Pre	S.D.	Post	S.D.		
Anglos (N - 11)	85.09	12.6	91.00	15.6	6.91	2.77 *
Latins (N - 8)	77.50	7.1	83.25	6.9	5.75	1.98
Negroes (N - 14)	81.29	11.6	88.00	11.6	6.71	3.05 **

** $p < .01$

TABLE 7

MEANS, STANDARD DEVIATIONS, AND COMPARISON OF DIFFERENCES FOR PRETEST AND POSTTEST SIT IQ SCORES FOR BOYS AND GIRLS IN DIFFERENT ETHNIC GROUPS WITHIN THE E GROUP

Group	Test Mean				Gain	t
	Pre	S.D.	Post	S.D.		
Anglo Boys (N - 5)	85.80	16.9	98.40	17.2	12.6	5.06 **
Anglo Girls (N - 6)	84.50	5.1	84.83	10.3	.33	.13
Latin Boys (N - 4)	73.75	4.5	79.75	2.9	6.0	2.06
Latin Girls (N - 4)	81.25	6.3	86.75	5.7	5.5	1.96
Negro Boys (N - 9)	83.00	13.7	92.77	10.6	9.8	4.45 **
Negro Girls (N - 5)	78.20	5.6	79.40	6.1	1.2	.54

** $p < .01$

gains while girls virtually stood still.

Four boys and four girls pretested on the opening day had had some previous schooling, it was discovered. They were given posttests along with the E subjects, but results were handled as a separate "prior schooling" group. Results of an analysis of variance, shown in Table 8, revealed that the sex x testing interaction was nonsignificant. However, the trends of this group appear quite different when their mean IQs, shown in Table 9, are compared with the E group means shown in Table 7. The eight children who had previous school experience, all of them Negro, had higher IQ means on the pretest than any of the E subgroups. Boys were 10 IQ points ahead of girls on the pretest, but the girls gained 8 points between tests, which was statistically significant (p .01) while the boys gained only 2 points, which was non-significant, as shown by t tests of direct differences.

During the posttesting of the C group it was learned that 10 of the 23 subjects had received some coaching from their families on items missed in the pretest. In order to assess the effect of this unexpected variable, the C subject data was divided into coached and noncoached categories, as shown in Table 10. A t test for direct differences indicated that the gain of the coached group was significant ($p < .01$)

TABLE 8

SUMMARY OF ANALYSIS OF VARIANCE OF PRETEST AND POSTTEST SIT IQ SCORES FOR BOYS AND GIRLS IN PRIOR SCHOOLING GROUP

Source	df	MS	F
Between Subjects			
Boys - Girls (B)	1	121.00	<1
Error (b)	6	233.50	
Within Subjects			
Pre - Post (A)	1	49.00	1.58
A X B	1	9.00	<1
Error (w)	6	31.00	

TABLE 9

MEANS, STANDARD DEVIATIONS, AND COMPARISON OF DIFFERENCES FOR PRETEST AND POSTTEST SIT IQ SCORES FOR BOYS AND GIRLS WITH PRIOR SCHOOLING

Group	Test Mean				Gain	t
	Pre	S.D.	Post	S.D.		
Boys (N - 4)	97.95	11.9	99.75	12.8	2.0	1.01
Girls (N - 4)	87.75	8.1	95.75	7.8	8.0	4.06 **

** $p < .01$

TABLE 10

MEANS, STANDARD DEVIATIONS, AND COMPARISON OF DIFFERENCES FOR PRETEST AND POSTTEST SIT IQ SCORES FOR COACHED AND NONCOACHED SUBJECTS IN THE C GROUP

Group	Test Mean				Gain	*t*
	Pre	S.D.	Post	S.D.		
Coached (N - 10; 2 M, 8 F)	95.4	12.1	102.4	10.3	7.0	4.73 **
Noncoached (N - 13; 7 M, 6 F)	87.6	14.7	90.7	12.7	3.1	2.09 *

* $p < .05$
** $p < .01$

and that the gain of the noncoached group was somewhat less significant ($p < .05$). As an analysis of variance was not run for coached and noncoached C subjects, a t test for uncorrelated means was made, which indicated that the mean difference in gain of 3.9 IQ points was nonsignificant.

The preceding data has shown that most, but not all, the groups of E and C subjects made significant gains in mean IQ within two or three weeks of the commencement of the Head Start program. However, the pattern of gains of boys and girls was different between E and C groups, and between E and a prior schooling group attending the same Head Start classes. Relative gains of boys and girls also differed in the various ethnic groups represented in the E group. These findings lead to the conclusion that pretest timing does make a difference in regard to evaluations of the impact of this type of program upon children, since only a very early measurement would sample what all the groups were like in relation to each other, prior to exposure to the educational intervention.

CHAPTER IV

DISCUSSION

This study was directed to the question of when a pretest should be administered to children in a compensatory education program such as Head Start, in order to facilitate accurate evaluation of the effects of the program. Along the way several other methodological questions were touched upon, including the feasibility of giving the pretest on the first day, how to secure and when to test the control group, the danger of practise effect due to repeated testings, and the danger of coaching of subjects. As data were examined, an unexpected significant difference emerged between scores of E boys and girls on the posttest. No conclusions can be reached concerning the cause of this difference, but certain possibilities are suggested in a later section.

Timing of the pretest

The question of proper pretest timing was investigated by administering pretests to the E subjects on the opening day of Head Start, and retesting them either two or three weeks later. If the E subjects had gained significantly more than the C subjects it would indicate that two weeks would be

too long to delay the administration of the pretest. Since the E subjects did not gain significantly more than the C subjects, it would seem at first blush to indicate that delayed pretesting would be all right. However, part of the E group, the boys, changed significantly more than the girls, while C boys and girls did not shift in relative IQ standing. Thus the boys in Head Start rapidly became different, and the relationships between groups of subjects was modified also. The shifting standings of boys and girls was observed in Anglo and Negro subgroups but not in the Latin ethnic group, further altering relationships. Children in Head Start who were known to have had a previous school experience showed a tendency to a trend opposite to the E group, with girls gaining more rapidly than boys. The over all effect of the different patterns of IQ changes within the first few weeks is to demonstrate that pretest timing would definitely affect the accuracy of later evaluations of the impact of Head Start type educational intervention programs. Two weeks has been demonstrated to be too long a delay. Further studies may be able to narrow down still further the range of acceptable pretest times which would still give a sufficiently accurate picture of the status of the subjects prior to treatment.

Several procedures in the pretest were tried out to

determine their workability. The pretests were administered on the opening day of the regular Head Start session, and this was found to be a workable procedure. A special team of examiners was brought in just for the pretests, since the regular staff were all fully occupied; this in itself had an advantage, as will be discussed under the topic of coaching. The regular staff were able to take time to implement the examiners' work, the teachers did not appear to be disturbed by the random selection and removal of subjects from class, and the children, though showing signs of general bewilderment, responded well to the individual tests.

Problems Connected with the Control Group

Difficulties were encountered in locating a suitable population of six year old children, unexposed to school, for a control group. Once sampled, the C group turned out to be somewhat different from the E group, and had significantly higher mean IQ scores on both pretest and posttest. This may have been a reflection of the apparent upward mobility of the general C population, plus the lack of an adequate Negro and Latin sample. This before-treatment difference in E and C subjects would have posed a serious problem if this study had been an actual evaluation of the impact of Head Start treatment. However, since the study was exploring a

methodological question, timing of a pretest useful for evaluation, the C group was deemed adequate since it did meet the chief criteria of age and lack of schooling.

Some differences in the test administrations given E and C groups were built into the experimental design, in order to solve the practical problems of arranging test schedule and location. The testing dates for the two groups were one week apart, in order to utilize the same team of examiners for both groups. The staggered schedule seemed to work out well, as the same hot weather and general conditions prevailed during both testing weeks, and subjects in each group averaged exactly the same chronological age, 6 yr. 3 mo.

The only difference that might have been caused by the staggered schedule would have been a greater amount of experience on the part of the examiners during the C test sets. One study done with Head Start children purported to find a difference in Stanford-Binet IQ scores which was related to examiner experience (Smith, May, & Lebovitz, 1966). On the other hand, four out of five studies of examiner experience, using subjects of various ages, which were reviewed by Sattler and Theye (1967) did not find experience to be a significant variable in accounting for IQ score differences. Smith, et al set the criterion for examiner experience at 20 test adminis-

trations; experienced female examiners gave significantly lower scores, experienced male examiners gave slightly higher ones. Since the present study utilized examiners of both sexes, and since it is considerably easier to learn to administer and score the SIT than the Stanford-Binet, it was concluded that examiner experience would not be likely to be a critical variable in this study.

The differences in testing situations between E and C groups did turn out to be an important variable, because it was found that the C subjects' homes were often not satisfactory places to conduct a test, especially since an interested group of onlookers usually observed the whole pretest. These parents and siblings were not warned against later coaching of the subjects, on the theory that such an instruction would do more to suppress the report of coaching than the activity. During the course of the posttest visit, nearly half the mothers volunteered remarks about drilling the subject on missed items. This modification of the C group would not have occurred if the C subjects could have been pretested in the same sort of privacy available for the E pretests.

Coaching

It is hoped that the experience of this study with the problem of C group coaching will werve as a warning to future

researchers about a hazard to avoid. In other, quieter summers it may be possible to dismiss onlookers without risking suspicion. Or perhaps the pretest could be given in the examiners' car, still within sight of the family but out of earshot. Some researchers have even been able to fit out minibusses as mobile testing laboratories, which would make comprehensive medical as well as psychological testing possible right at the subjects' homes.

The discovery of coaching suggests further that it would be an error to think of the C condition as a "no treatment" condition. The C subjects are learning at home at the same time that the E subjects are receiving the formal educational treatment of Head Start. Yet the C condition is the same homeenvironment which, in the case of the E children, was supposed to be so lacking that the remedial intervention of Head Start was needed. Studies need to be made of the latent educational possibilities at home before the merit of an early schooling program such as Head Start can be fully decided. The DARCEE project (Miller, Forrester, Gilmer, & Cupp, 1966) has begun to report research with groups of subjects who receive special education at home from their mothers, who have in turn been trained by home visitors.

Coaching, it should be noted, could alter the results

of the E group just as well as the C group, if the pretests are given by those who later teach the E group, since there would be a natural inclination to teach toward mastery of those areas in which the children had scored low. For this reason teachers as well as mothers should not be aware of the specific nature of the evaluative instruments used to assess a treatment, unless their coaching is to be considered as a part of that treatment.

Practice Effect

A danger inherent in the pretest - posttest research design is that practice on the pretest may artificially raise subjects' posttest scores. A slightly different sort of practice effect was noted by Campbell (1957) who found that adults given an attitude measure as a pretest were alerted as to what they should be learning during the experimental treatment. McBeath (1965) found a practice effect operating with above-average children who were pretested and then trained on a perceptual task. However, Entwisle (1961) found that although the pretest interacted with the IQ scores of the high scoring children, it did not affect the IQ scores of the average or below average children in the sample.

Since Head Start children typically test in the below average IQ range, it might be assumed that the pretest would

produce very little practice effect. Data from the present study and from two other studies which included some of the E subjects, indicate that there is but little practice effect due to repeated testing of Head Start children with the SIT.

The SIT manual (Slosson, 1963, p.v) provides the information that the average test-retest gain was 2.3 IQ points, with a two month test interval. This is the average practice over all age groups and capability levels. It was expected that both E and C groups in the present study would gain at least that much, especially since the posttest was in only three weeks. Both groups as a whole gained a good deal more than that; however, E girls gained less, and the noncoached C subjects gained only .8 points more than the predicted practice effect.

During the second week of Head Start at Aldine another research project was being conducted, which involved giving the SIT and two other tests to 155 subjects, for later comparison with the children's school grades (Hutton, G., personal communication). Mean IQs by sex and ethnic group are shown in Table 11, and should be compared with E data in Table 7. The Hutton sample does contain 11 E subjects who were inadvertantly included, but all the rest had had no practice on the SIT. The mean IQs of the Hutton sample and the E group

TABLE 11

MEANS AND STANDARD DEVIATIONS OF SIT IQ SCORES FOR BOYS AND GIRLS IN DIFFERENT ETHNIC GROUPS TESTED DURING SECOND WEEK OF HEAD START

Group	Mean IQ	S.D.
Anglo Boys (N - 16)	97.12	10.0
Anglo Girls (N - 19)	87.63	16.4
Latin Boys (N - 20)	84.65	13.1
Latin Girls (N - 24)	84.58	14.7
Negro Boys (N - 42)	90.05	14.2
Negro Girls (N - 34)	88.79	14.8

posttest mean IQs are very close for each sex x ethnic group, indicating that these variables plus the effects of a few days in Head Start are more significant influences on the IQ level than practice effect from a pretest.

Another even larger administration of the SIT took place from the fourth - eighth weeks of Head Start at Aldine, in which the Head Start staff and teachers gave the SIT to every pupil remaining untested up to that time, so that the test could be used for guidance and referral for all the children (Sammons, G., personal communication). No less than 18 of the E subjects were unintentionally included and thus took the SIT for a third time. If practice on a test were ever to make a difference on the score, it would surely have shown up in higher scores for these children, compared with the rest of the Sammons sample. However, their IQ means were exactly like those of the rest, as shown in Table 12. The nine boys gained more than the predicted practice effect on the second test, then appeared to lose somewhat on the third test. The nine E girls gained less than the predicted practice effect on their second test, but gained more than that on the third.

The cumulative evidence all indicates that practice coming from the pretest is not of enough influence to present a problem for research design for Head Start evaluations.

TABLE 12

PRETEST, POSTTEST, AND THIRD TEST MEAN SIT IQ SCORES FOR 18 E SUBJECTS, AND MEAN SIT IQ SCORES OF CHILDREN TESTED ONCE DURING FOURTH - EIGHTH WEEKS OF HEAD START

Group	Pre	Post	Test During 4th-8th Week
E Boys (N - 9)	79.44	87.77	85.85
Other H.S. Boys (N - 126)			85.96
E Girls (N - 9)	81.55	82.22	87.22
Other H.S. Girls (N - 130)			87.15

Serendipity

The data, collected for study of the timing of the pretest, contained a surprise - a pronounced and statistically significant difference in the mean IQ gains of boys and girls in their first few weeks' encounter with classroom education. The boys made significant gains, while the girls gained even less than the expected practice effect. The remainder of this study will be a consideration of a number of plausible hypotheses which might, or might not, explain this serendipitous finding.

Hypotheses Concerning the Sex-Linked Difference in IQ Gains

1. Biased sample. The most likely explanation of the sex-linked difference in IQ gains would be that the small E sample, although randomly selected, was biased in some way in favor of boys. Such a theory would have to explain away the equality of the sexes on the pretest, and also the expectation, based on girls' greater maturation at age six, that the girls would have higher IQ means than the boys. Girls did exceed boys in the C sample. Furthermore, the large fourth - eighth week testing, when analyzed by sex and ethnic group as shown in Table 13, placed girls above boys in the Anglo and Negro groups, though not in the smaller Latin group.

However, the IQ means for the 155 children tested during

TABLE 13

MEANS AND STANDARD DEVIATIONS OF SIT IQ SCORES FOR BOYS AND GIRLS IN DIFFERENT ETHNIC GROUPS TESTED DURING FOURTH - EIGHTH WEEK OF HEAD START

Group	Mean IQ	S.D.
Anglo Boys (N - 42)	89.97	15.2
Anglo Girls (N - 51)	93.26	11.9
Latin Boys (N - 22)	84.04	14.3
Latin Girls (N - 9)	73.22	12.7
Negro Boys (N - 73)	84.71	10.9
Negro Girls (N - 84)	85.15	8.2

the second week of Head Start (see Table 11) show boys consistently ahead of girls. The ethnic x sex means parallel those of the E group posttests (see Table 7). The larger sample included about one third of the total group of Head Start pupils; it is good evidence that the E posttests were reflecting the actual situation between boys and girls. Furthermore, the E subjects tested a third time (see Table 12) were comparable to the rest of the E group on pre and post-test means (see Table 3), and were also representative of the 281 children tested during the fourth - eighth weeks. If one accepts the representativeness of these 18 subjects, it leads to the possible theory that girls' IQ levels respond little to the first few weeks of Head Start and then go up, while boys' IQ level rise considerably and at once, then quickly reach a plateau and perhaps even decline a bit, as education continues.

Another study of children in their first encounter with formal education has actually found the same phenomenon. McNeil (1964) pretested kindergarten boys and girls, finding them equal on a reading readiness test, then trained them for three weeks with programmed reading instruction machines, then posttested them. The boys scored significantly higher. The same subjects were retested during first grade and girls now scored significantly better. McNeil, after considering several

possibilities, suggested that his shifting sex difference might be due to a combination of two factors: first grade teachers perceive boys more negatively than girls and give them less opportunity to practice reading, while the teaching machine was impartial; and the individualized programmed instruction reduced peer group activity, which increased the boys' attentiveness to the learning task and brought out a better than usual performance from them.

2. Teacher expectations. McNeil (1964) blamed the first grade teacher's negative attitudes, in part, for the boys' worsened performance. Likewise, Rosenthal and Jacobson (1968) found that disadvantaged children whom teachers expected to be brighter, due to spurious IQ data, actually gained significantly more in IQ than children about whom no such expectations had been aroused. However, this hypothesis would not explain the shifting sex differences of the E sample unless it could be shown that something changed teachers' expectations for boys and girls shortly after the third week of Head Start.

3. Ethnic differences. Tables 7 and 11 both show the patterns of sex x ethnic differences for Aldine Head Start children, indicating that boys exceeded girls, during the early weeks of Head Start, among both Anglo and Negro ethnic groups,

but not of the Latin group. Thus the difference was not confined to one particular ethnic group within the total sample.

4. Examiner characteristics. A number of studies, such as those reviewed by Sattler and Theye (1967) have found interactions between subject and examiner characteristics. However, preliminary analysis of the data of this study revealed no trends toward interaction between (a) sex of examiner and sex of subject, and (b) experience of examiner and sex of subject.

5. Amount of attendance. Since Berzonsky (1967) noted a correlation between IQ gains and number of days attendance at Head Start, this possibility was also checked. Perhaps boys had been more regular in attendance than girls during the first weeks of Head Start. Attendance records revealed that almost all the E subjects had perfect attendance through the first three weeks, perhaps a result of the door-to-door but service, so that no correspondence could be found for attendance and IQ gains.

6. Vulneribility to Environment. In a review of the findings of the Berkeley Growth Studies, Bayley (1966) found that a hostile early environment has a lasting impact on boys' IQs, but not on girls' IQs. She concludes (p. 106),"Boys appear less able than girls to recover from hostile, rejecting treatment; but they may also profit more, in the long run, from understanding, loving acceptance." Perhaps this explains the relatively larger gains of the boys so soon after they entered

friendly, accepting atmosphere of the Head Start classes, while girls, by nature more independent of the environment, were less affected by the positive change.

Another scrap of evidence comes from a study done by Peterson and reported by Unikel (1967), in which Head Start boys were found to perform slightly better than girls under the condition of "praise of person" while the girls did better under "praise of performance", whereas middle class preschool boys and girls both did better under the praise of performance condition. Perhaps Head Start teachers provided a praise of person environment to such an extent that boys were encouraged far more than girls among the E sample.

Though the above theory of environmental vulnerability might explain the initial rapid gains of the boys, it would not account for the data indicating this advantage disappeared in a few weeks, since the Head Start environment contained few if any changes over the eight weeks.

7. Attentional Deficits. Kohlberg (1967) did a study of attentiveness among a group of disadvantaged preschoolers in a Head Start program, contrasted with a group of middle class children in a Montessori program. He contended that attention, next to IQ, is the most powerful predictor of school learning, and that attentional deficits are pronounced among disadvantaged

children. This may be due to the overcrowded environment of the disadvantaged, where there is very seldom an opportunity to become engrossed for long in something interesting without interruption from others. A permissive classroom, with few rules and much group process, would do very little to help a child make up for attentional deficits. Kohlberg actually noted a decline both in measured attentiveness and IQ among the Head Start children in the permissive classroom. Unfortunately he did not divide his data by sex. McNeil (1964) had also blamed part of his boy subjects' decline upon lack of attentiveness in first grade. Perhaps the nine E boys tested three times showed a slight IQ decline for the same reason.

If, as Bayley (1966) says, girls are less vulnerable to the environment, it would follow that disadvantaged girls would show less of an attentional deficit than boys. This would account for many of the findings cited, such as the girls' greater resemblance to middle class children, girls' superiority over the long run in both Head Start and first grade, and even the first grade teachers' more favorable opinion of the girls. However, it would not account for one last puzzling element, the virtual standstill in IQ level of the E girls during the first weeks of Head Start.

8. Readiness for leaving home. Perhaps, for a final hypothesis, the shock of leaving home for school for the first

time is greater for disadvantaged girls than it is for boys. Boys of all ages, in a lower class neighborhood, seem to have more freedom to roam away from home than do girls, which may be the reason more girls than boys were found at home for the C sample, and more girls than boys received coaching from their mothers. If this informal observation could be supported with data it might provide an explanation for the initial standstill of the E girls, who were undergoing a sort of weaning shock in the first weeks away from home.

The previous eight hypotheses were all suggested by the intriguing finding of shifting sex differences in the E sample, with the last three theories appearing to have most merit. More studies need to take a close look at what happens to children in their first few weeks in Head Start and other forms of education, with eyes open for the possibility of significant sex differences. This is not being done currently. For instance, an excellent study just published by Zigler and Butterfield (1968) of children in a year-long Head Start, did pretests and then retested subjects three weeks later, and then twice again at the end of the year, and reported no sex data. Only when numbers of studies have reported sex difference data, will it be possible to decide whether a shifting sex difference is a frequent phenomenon, and to determine cause or causes.

CHAPTER V

SUMMARY

This study explored a problem encountered in evaluating Head Start programs in the field, the question of the timing of the pretest. The opening day of the Head Start summer session was selected for study, as the time most likely to be feasible in practice and yet yield valid pretest data. A random sample of children was chosen and given the Slosson Intelligence Test on the first day of Head Start. The subjects were posttested two to three weeks later and were found to have gained significantly in IQ scores. The E boys gained significantly more than the E girls. Anglo and Negro boys gained more than counterpart girls, while Latin boys did not.

A C sample of children the same age, who were not yet in school, were given the same pretest in their homes one week later, and also were given the same posttest the week after the E posttests. The C subjects also gained significantly in mean IQ, but relative positions of boys and girls did not change. Evidence was found that posttest scores of some C subjects may have been raised by parental coaching between tests.

Several conclusions may be drawn from analysis of data

from E and C groups, which are strengthened by reference to other, larger samples from the same Head Start population.

1) Pretesting on opening day, which proved to be workable in practice, is preferable to any later date, since the early weeks of school appear to affect the sexes differently. The pretest should not be delayed beyond the second week. This study was not arranged to narrow down the cut-off time with more precision than this.

2) Subjects in the C group were tested in their homes, with parents often present, with the result that some C subjects were coached on missed test items before the posttest. Testing conditions for the C subjects need to be arranged to prevent such coaching, if E and C treatments are to be properly compared.

3) Practice effect resulting from repeated administrations of the SIT to E subjects was demonstrated not to be a problem among Head Start children.

4) Shifting sex differences were found when analyzing the data. Boys gained significantly more than girls during the first few weeks, but evidence indicated that this advantage disappeared in later weeks of Head Start. A number of possible explanations of this phenomenon were considered and discarded. The E sample did not seem to be biased, but was demonstrated

to be representative of the sex and ethnic groups in the total Head Start population of the school district. Other possible influences were also discarded: teacher expectations, ethnic differences, examiner differences, or days of attendance.

Three possible causes were found which, alone or in some combination, could account for the sex-linked differences found by this study: boys may be more responsive than girls to a positive environment such as Head Start, boys may suffer from greater deficit in attentiveness than girls, and finally, the girls may be more home-oriented than boys when they first enter school, so that they require a longer adjustment period before they can begin to benefit from school learnings. It was suggested that further study of differences in boys' and girls' responses to initial educational experiences would be fruitful, and should be added to Head Start evaluations being done currently.

BIBLIOGRAPHY

BIBLIOGRAPHY

Alpern, G. D. The failure of a nursery school enrichment program for culturally disadvantaged children. American Journal of Orthopsychiatry, 1966, 36, 244-245.

Bayley, N. Developmental problems of the mentally retarded child. In I. Philips (Ed.), Prevention and treatment of mental retardation. New York: Basic Books, 1966. Pp. 85-110.

Beller, K., & Nash, A. Research with educationally disadvantaged pre-school children. Paper presented at the American Educational Research Association Annual Meeting, Chicago, February, 1965.

Berzonsky, M.D. An evaluation of an eight-week Head Start program. Paper presented at the American Educational Research Association Annual Meeting, New York, February, 1967.

Campbell, D. T. Factors relevant to the validity of experiments in social settings. Psychological Bulletin, 1957, 54, 297-312.

Eisenberg, L., & Connors, C. K. The effect of Head Start on developmental processes. Paper presented at the Joseph P. Kennedy Jr. Foundation Scientific Symposium on Mental Retardation, Boston, April, 1966.

Entwisle, D. R. Interactive effects of pretesting. Educational and Psychological Measurement, 1961, 21, 607-620.

Giles, D. E., & Daniel, A. A. The effects of Project Head Start activities on the oral language of first-grade pupils. Paper presented at the American Educational Research Association Annual Meeting, New York, February, 1967.

Gray, S., & Klaus, R. A. An experimental preschool program for culturally deprived children. Child Development, 1965, 36, 887-898.

Guilford, J. P. Fundamental statistics in psychology and education. New York; McGraw-Hill, 1965.

Kohlberg, L. Montessori with the culturally disadvantaged: a cognitive-developmental interpretation and some research findings. In R. Hess & R. Bear (Ed.), Early education: Theory, research, and action. Chicago: Aldine Press, 1967.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 195[illegible].

McBeath, M. The effect of pretest interaction on a perceptual task. California Journal of Educational Research, 1965, 16, 203-209.

McNeil, J. D. Programmed instruction versus usual classroom procedure in teaching boys to read. American Educational Research Journal, 1964, 1, 113-119.

Miller, J. O., Forrester, B. J., Gilmer, B., & Cupp, J. A study of older siblings of children in a preschool intervention project. DARCEE Papers and Reports, 1966, 1 (2). Nashville: Demonstration and Research Center for Early Education, Peabody College.

Rosenthal, R., & Jacobson, L. R. Teacher expectations for the disadvantaged. Scientific American, 1968, 218, April, 19-23.

Sattler, J., & Theye, F. Procedural, situational, and interpersonal variables in individual intelligence testing. Psychological Bulletin, 1967, 5, 347-360.

Slosson, R. L. Slosson Intelligence Test (SIT) for children and adults, 1963 ed. East Aurora, N.Y.: Slosson Educational Publications.

Smith, H. W., May, W.T., & Lebovitz, L. Testing experience and Stanford-Binet scores. Journal of Educational Measurement, 1966, 3, 229-233.

Tannenbaum, A. J. An early intervention program that failed. Unpublished manuscript, Teachers College, Columbia University, 1966.

Unikel, I. P. Alternation behavior by children from lower socio-economic status groups. Paper presented at the Southwestern Psychological Association Meeting, Houston, 1967.

Weikart, D. P. (Ed.) Preschool intervention: A preliminary report of the Perry Preschool Project. Ann Arbor: Campus Publishers, 1966.

Wolff, M., & Stein, A. Head Start six months later. Phi Delta Kappan, 1967, 48, 349-350.

Zigler, E., & Butterfield, E. C. Motivational aspects of changes in IQ test performance of culturally deprived nursery school children. Child Development, March 1968, 39 (1), 1-14.