DOCUMENT RESUME

ED 130 229 CS 002 956

AUTHOR Palmer, Francis H.
TITLE The Effects of Minimal Early Intervention on Subsequent IQ Scores and Reading Achievement. Final Report.
INSTITUTION State Univ. of New York, Stony Brook.
SPONS AGENCY Education Commission of the States, Denver, Colo.
PUB DATE 76
CONTRACT 13-76-06846
NOTE 63p.

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS Early Childhood Education; Elementary Education; Followup Studies; *Intelligence Quotient; *Intervention; Lower Class Males; Negro Youth; *Preschool Evaluation; Preschool Learning; *Preschool Programs; Program Evaluation; *Reading Achievement; *Success Factors

ABSTRACT

IQ and reading achievement in grade five were examined in a ten-year follow-up study of children who had participated in an early-intervention program, at ages 24 or 36 months. The intervention program varied age of training, type of training (concept versus discovery), and social class for 310 black male children from Harlem. The follow-up study obtained WISC scores for 139 and reading scores for 117 of the original sample. Analyses indicated these were representative of the original experimental and control samples. Comparison groups not involved in the original study were also drawn. Results indicated that concept training at age 24 months or 36 months significantly affected reading in the fifth grade and IQ at ages 10 to 12. Intervention at age two had an effect on reading and IQ, whereas intervention at age three affected IQ but not reading. Discovery training affected IQ but did not affect reading. Implications of the findings for general evaluations of the success or failure of Headstart and other early-intervention programs are discussed. (AA)

FINAL REPORT

THE EFFECTS OF MINIMAL EARLY INTERVENTION ON SUBSEQUENT IQ SCORES AND READING ACHIEVEMENT

By

Francis H. Palmer

State University of New York, Stony Brook

Follow-up Research done in Contract (13-76-06846) with the Education Commission of the States over the time period October 1, 1975 to August 31, 1976.

FOREWORD

August 22, 1976

There have been many critics of the effects of early intervention. Critics who claimed that Black children could not be affected because they are genetically inferior and unequipped to benefit from the public schools as their middle class peers do. Critics who claimed that the number of hours a child is exposed to intervention is insufficient to counteract the devastating effects of the ghetto family and community.

These data suggest that those critics are wrong.

When the funds were made available by ECS for this follow-up study, this investigator did not hope that the results would be as convincing as he now thinks they are. The study shows an intervention at a specific period of time had a clear and meaningful effect on the subsequent scholastic achievement of both middle and lower-class Black boys.

Furthermore, this study provides information about two types of training, and two times when training was introduced, and the effects those variables have on subsequent school performance. This investigator is too close to the data at this time to speculate on how these results shall ultimately modify the consensus of belief about age and type of training - but the data are clearly relevant to those compelling subjects.

If one wants to know how early childhood intervention affects scholastic performance at age 12, someone must support a follow-up to assess children at age 12. ECS provided that support.

Surely, having the opportunity to obtain these results is more personally rewarding than had I invented the transistor.

THE EFFECTS OF EARLY INTERVENTION ON IQ AND READING ACHIEVEMENT

I. PURPOSE

The purpose of this study was to locate and assess in the fifth grade as many of the original 310 children in the Harlem Research Center early intervention program as possible. That study, initiated in 1966 with a grant from the National Institutes of Health, provided early intellective training for 240 two and three-year-old Black male children, as well as annual assessment of 70 control children who did not participate in the training program. At grade 5, achievement is defined as the IQ derived from the Wechsler Intelligence Scale for Children (WISC), and the reading achievement scores obtained from the school each child attended.

II. BACKGROUND

A. Research Design.

The study varied age of training, type of training, and social class.

The age at which training began in 1966 was 24 months (T-2) and in 1967 36 months (T-3).

At each age, 120 children served as subjects. (T-2=120; T-3=120)

Type of training consisted of two conditions: Concept Training (CT) and Discovery Training (DT). Both groups attended the same facility twice weekly for one hour over a period of eight months. Both were exposed to one-to-one instruction, the same instructors, and identical training materials. Instructors were rotated every six sessions. No child had the same instructor more than six. Conditions were identical except for what occurred during the training session. The CT Group received a structured curriculum designed to teach concepts such as big-little, rough-smooth, wet-dry, loud-not loud, next to-far away, etc. Those concepts were taught

in a fixed procedure involving four steps with a specified criteria for knowing each concept. The instructor planned for each session, predetermining what concepts would be taught, and with what materials the concept would be taught. He guided the training, initiated most conversations, and recorded the results of each session so that he or subsequent instructors knew what had occurred to the child previously. The DT group had no curriculum nor any structured training. Instructors did not initiate conversations, but responded to questions and gestures the child made. Instructors were trained to not emphasize the concepts being taught in the CT group. In many respects, the training was passive; not unlike what occurs in many child care centers except that the one-to-one situation prevailed at all times. Detailed descriptions of CT and DT conditions have been documented before. (Palmer, 1971)

Studies were conducted during the training period to ascertain that instructors differentiated between their CT and DT roles. Those studies showed, for example, that the DT training involved only 12% as much conversation with the child as the CT training did. Thus, it was shown that the instructors provided the two types of training specified by the research design.

Social Class was defined by the education and occupation of the parents, and measured by the Hollingshead-Redlich Two Factor Scale of Social Class, slightly modified. Those children designated Lower Class (LC) in this study comprised Category V of the Hollingshead-Redlich. Those children designated Middle Class (MC) were of parents who were categorized IV to II on that measure.

The original sample was representative of the larger Harlem population of 1966. Of those ultimately classified as lower class (LC) 38 were semi-skilled production workers and the remainder were unskilled, unemployed, or on welfare. Twenty-three percent (23%) had completed high school, and 34% had never attended high school. The MC parents were comprised of 75% in the Hollingshead Category IV---clerical personnel, skilled laborers, machine operators, etc.; 25% were in Categories III, II, and I---executives, administrators, business managers, etc. Eighty-four percent (84%) of MC had graduated from high school, and twelve percent (12%) of that subset had graduated from college.

B. Subject Selection

The 310 subjects in the original sample were selected from 1,500 birth records of black male children born in the Harlem and Sydenham Hospitals in Manhattan between August and December of 1964. Each lived between 100th and 145th Streets in Manhattan at that time. The information on the 1,500 birth records was sent to the U. S. Post Office, who confirmed 900 current addresses from which 700 home interviews were attempted and 500 were achieved. From that 500, the subject pool was established with those children who met the following criteria: over five pounds at birth, mother with no history of syphilis or drug addiction, both parents self-described as Negro, both parents spoke English as a first language, and no serious illnesses between birth and 24 months of age. The 310 subjects were drawn from that pool to meet a predetermined distribution by social class. That distribution specified more lower class in each relevant cell than middle class because it was assumed attrition would be greatest for the former.

Analysis of the selection process showed no bias across those assigned to the several cells of the design at either age of training.

The percentages located, interviewed, and if interviewed, the proportion who rejected the program did not differ by group after assignment was made.

Ultimate assignment to cells of the design was made on the basis of an interaction between MC and LC, and on pre-training assessment using the Concept Familiarity Index (CFI), a measure of already existing knowledge about the same concepts included in the CT curriculum. CT, DT and Control, by age of training, did not differ on means and standard on the CFI, nor on social class.

C. The Training Experience

The T-2 group was trained in 1966, the T-3 group in 1967. The training facility included eight 9' x 7' rooms with one-way mirrors and audio equipment which provided for observation. Parents and children were provided transportation, if desired. Each child had regular appointments every week for two one-hour sessions, staggered with at least one day intervening the last. Parents were encouraged to attend at least the first six sessions, to be present while the child adapted to the Center, and to provide the opportunity for them to see and hear what activities their child would be engaged in. After the first six sessions, parents could attend when they desired, and if anything occurred in the training session they wanted explained, they were free to enquire of instructors and other staff personnel.

No parent observed CT training if their child was given DT training, and vice versa. None were remunerated except in the form of transportation.

Over the two years of training 82% of all appointments were met. Each child that completed training had at least 40 sessions, and the average child received 45. In addition, the average time required for pre-training assessment was eight hours for T-2 and six for T-3. The average post-training assessment time for training groups of both ages was six. Of the 120 two-year-olds who began training in 1966 (T-2) 100 completed training eight months later. Of the 120 three-year-olds who began in 1967, 114 completed the program.

D. Results of Preschool Assessments

Participating children (T-2 and T-3) and the controls were assessed when training ended and annually thereafter until they were four years and eight months old. Consequently, those trained at two and the controls were assessed at 2/8, 3/8 and 4/8; and those trained at three were assessed at 3/8 and 4/8.

Although T-2, T-3 and the Controls were randomly assigned from the subject pool, a difference of 5 points on the Stanford Binet IQ was found between the controls at 2/8 and the T-3 group in post training assessment at age 3/0. For this reason, subsequent analyses were performed on the data with covariate techniques using IQ at first assessment as the covariate.

Assessment at 2/8 showed that T-2 outperformed the Controls on 14 of 17 behavioral measures. No differences existed between CT and DT groups.

Assessment at 3/8, when T-3 had just completed training and T-2 had completed one year before, showed that both groups were superior to the controls on most measures in the 14 test battery administered. No differences existed between CT and DT.

At age 4/8, most of the differences previously found between T-2

and T-3 and their controls disappeared. A statistically significant difference remained for performance on the battery as a whole, across measures, but only one individual measure differentiated those who had been trained and the Controls. Again, no differences existed between CT and DT groups.

With respect to performance as a function of the social class of the subject, differences between LC and MC did not exist consistently at 2/8 and 3/8, but at 4/8 they were clearly established (Palmer, 1970).

The results of the early assessments indicated that early intervention made immediate effects which were undifferentiated as far as age of training and type of training were concerned, but that effects were barely distinguishable at age 4/8. However, measures taken at 4/8 are not, cannot be, measures of scholastic achievement. One cannot measure scholastic achievement until children attend school. Furthermore, measures of scholastic achievement in the first two grades tend to be unreliable. No reading scores are recorded in the NYC public schools in grade 1, and those recorded for grade 2 have a built-in floor effect which makes the value of the measure questionable. Beginning at grade 3, although the floor effect persists, measures of reading are reliable indices of the child's performance, and tend to become more reliable with age.

The question to be answered by this follow-up study of the Harlem sample is: What are the effects of age of training, type of training and social class on scholastic achievement? Scholastic achievement has been defined as reading in the third, fourth and fifth grades, and by IQ.

III. SUBJECTS LOCATED AND ASSESSED IN 1975

A. Subjects Found

Two hundred ninety two children of the original 310 completed training in 1967 and 1968 and were assessed, or were assessed as controls in 1967. That figure (N-292) represents the maximum possible subject pool for the follow-up study. One hundred thirty nine (48%) of that pool have been located and assessed on the Wechsler IQ as of August 22, 1976. One hundred seventeen (117) reading scores for the fifth grade (1975) were obtained by then. More IQ's than reading scores are available because a subset of those administered the WISC were absent when the reading tests were administered in the schools, or because the reading scores have not yet been found. (Most New York City elementary schools were closed during the summer of 1976.)

The number of children in each cell of the design at the time of the first assessment, the number found for assessment in 1976, and the percentage found of the maximum possible, is shown in Table 1.

For main effects, the following percentages of subjects were assessed: T-2=54%, T-3=51%, Control= 40%, CT=47%, DT=53%, LC=49%, and MC=52%. Except for the controls, each cell is represented by about the same percentage found; remarkable since no particular attempt was made to search for subjects in any particular cell.

As of this writing, the study has been funded for another year of follow-up. At the end of next year, it is reasonable to assume that the total subjects found will approach 200, a figure better than original estimates of attrition, made when the study was designed.

When the study was designed the subjects per cell desired for multivariate analysis was 15. More LC than MC were included in the original and more controls than experimentals because disproportionate

attrition was anticipated among LC and controls. For the controls, that estimate was well founded and of the original 68, 27 are found, three less than the design anticipated. LC, however, has remained in the program more than expected as compared to MC (49% vs. 52%), leaving the most serious cell deficiencies for multivariate analysis in MC x T-2 x (CT\DT).

Despite the accuracy of the original estimates for attrition, subsequent years of follow-up should locate every subject possible. Our present estimate of subjects which will be located eventually, is 67% - an increase of 50 subjects over initial projections.

For the present analysis, however, some cell sizes are inadequate for multivariate analysis. Thus, the present analysis is limited to main effects, using Chi Square and t-ratio statistics.

Are the subjects found representative of the original sample? Are they representative of Harlem fifth grade boys in 1975?

We conclude that the subjects found are representative of the original sample. Within the original LC sample, 44% of the Controls, 41% of the T-2's, and 47% of the T-3's have been located and assessed. Within the original MC sample, 53% of the Controls, 46% of the T-2's and 60% of the T-3's have been located and assessed. Slightly more MC's than LC's, and slightly more T-3's than T-2's have been assessed, but those proportions are not sufficient to bias the analysis. Within the MC sample, more of the group as a whole has been found but the dropouts have been disproportionately in Categories I, II, and III. For whatever reason, only 12 of an original 37 (32%) originally classified in those advantaged groups have been located.

The 139 subjects assessed in 1976 include 66 originally classified

TABLE I

SUBJECTS ASSESSED IN 1967 and 1968, SUBJECTS FOUND IN 1975, BY TREATMENT

	LOWER CLASS (LC)			MIDDLE CLASS (MC)		
Trained At 2(T-2)	Assessed '67-68	Found 1976	Percent	Assessed '67-68	Found 1976	Percent
CT	33	17	52	23	11	48
DT	31	14	45	23	12	52
Trained At 3(T-3)						
CT	30	12	40	28	14	50
DT	31	18	58	25	14	56
Controls						
CT	36	13	36	32	14	44
DT	161	74	46	131	65	50

in category V (47%), 61 classified in Category IV (44%) and 12 in Categories I, II, III (9%). That balance is believed to be representative of the fifth grade boys in Harlem in 1975.

B. Attrition

Are the subjects found for this analysis representative of the original sample on Intelligence Scores and Social Class?

To determine whether or not the sample found is biased, those found were compared with those not found on the last appropriate assessment with the Wechsler Preschool Scale (WPPSI) and/or the Stanford Binet. The results of that analysis show that the Controls found did not differ from those not found, T-2 found did not differ from T-2 not found, T-3 found did not differ from T-3 not found, CT found did not differ from CT not found, DT found did not differ from DT not found. We conclude that the sample available for this analysis is not biased with respect to previous scores on intelligence tests.

The same analysis by group was conducted for Social Class. Original scores on the Hollingshead-Redlich were compared with t-ratios between the sample found and not found. Appendix A shows no significant difference between those found and not found by group in the research design and by Social Class. We conclude that no bias exists with respect to the Social Class of those children found and not found.

C. The Addition of a Comparison Group.

Frequently, when longitudinal studies include a control group, the controls as well as the treatment groups show an effect. This phenomena has been attributed to two factors: (a) the fact that the control group

is participating in a study, the Hawthorne effect, and (b) the control group is assessed at the same intervals as the treated groups, providing them with a degree of test wiseness not found in naive samples.

With respect to the Hawthorne effect, we have only anecdotal evidence. Conversations with the Control mothers during the 1976 assessment indicated that because their children had been assessed annually when they were 2/8, 3/8, and 4/8; and again in 1976, some of those mothers considered their children a part of the Harlem program as much as mothers whose children who were trained did.

With respect to test wiseness, the Control group had on the average, 20 hours of experience with examiners and tests at ages 2/8, 3/8 and 4/8; presumably twenty hours more than children not involved in the program.

For those reasons, it was considered essential to obtain another Control group (a comparison Group) who had no previous relationship with the study. Ideally, that group would be drawn from the same population as those who had participated in the program.

To date, this has been possible for reading scores only. To obtain a comparison group for IQ, a selected sample would have to be identified, located, and assessed on the WISC - too expensive for our budget.

For the results reported below, the Comparison Group has been derived in two ways, the former more precise than the latter.

The Comparison Group for the fifth grade reading scores was obtained in the following manner:

(1) The public schools (N=15) in Manhattan, whose students were 90% or more Black were identified.

(2) The number of children in our sample and assessed in 1976, at-

tending each of those 15 schools was tabulated.

(3) The percentage of children reading at grade level or better in each of those schools was obtained.

(4) A weighted score was derived for each school by multiplying the number of our subjects in that school by the schools percentage reading at grade level, and a mean percentage reading at grade level or above was calculated for the weighted distribution of school scores.

(5) It was ascertained that the weighted measure was identical to the percentage in the school district which 50% of our sample attended (District 5, Manhattan).

(6) The assumption was made that District 5, which as a District is over 90% Black, was most representative of the children in our sample.

(7) A random sample of 5th grade reading scores was obtained from different schools and classes in that district, 100 boys and 100 girls. Girls read significantly better than boys.

(8) From 2,060 District 5 reading scores in the 5th grade (1975), 852 scores were obtained for boys only. Only those names for which there was no question of gender were used.

(9) The resulting distribution of scores for 852 boys, 90% of whom are presumably Black, was used in the analysis as that distribution most representative of our sample. We refer to that group as the Comparison Group for Grade 5 Reading results.

The Comparison Groups for Grade three and four are not so precisely derived at this time. The Comparison Group reported for those grades is derived from the distributions of scores representing both boys and girls in District 5 in 1975. Since the average boy is four months (5.0 vs. 5.4)

behind the reading level of the average girl in the fifth grade, we assume that the Comparison Group distribution used at grades three and four is a conservative estimate of the differences between our sample, all boys, and the sample used as representative of their peers.

IV. READING RESULTS

Reading results from the 1976 assessment include scores obtained when the subject population was in the third grade (1973) the fourth grade (1974) and the fifth grade (1975). They are presented below as they are a function of age of training, type of training, and social class - the variables manipulated in the research design - and by grade.

The results obtained include scores from over 50 elementary schools in the New York City public school system and in 15 schools in the Catholic diocese in New York City, as well as nine scores for children who have moved out of the City and now live elsewhere. Only scores directly comparable to the measures used in the public school system were used in the analysis.

In 1975, the New York Public Schools began using the Stanford Achievement Test (SAT) to assess the achievement of all pupils. In 1973 and 1974, and for several years previous to that, the Metropolitan Achievement Test (MAT) was used. So far as can be determined, the reasons for shifting measures were two: (1) the norms for the SAT were more representative of the children in the NYC public schools; that is, on the average their students would score higher on the SAT than the MAT, and (2) it was suspected that coaching for the MAT was rampant - the MAT had been used long enough so that items on that test were available. For the second of these reasons, the individual scores obtained for grade 5 are probably more valid

than for scores at grades four and three. Since the SAT was used for the first time in 1975, teachers had little opportunity to coach for it.

The shift from the MAT to the SAT explains the spurt in reading scores from grade four to five, as compared to the change from three to four. For Comparison Groups comprised of both boys and girls at each grade level in District Five, the average gain from the third grade to the fourth was only 4.4 months, whereas from grade four to grade five, the average gain was 1 7 months. Either fifth grade teachers are performing miracles or the SAT and MAT norms differ.

There are some 15 scores in the distribution obtained from those Catholic schools where the Stanford Reading Achievement Test (SRA) was used. Where students in the Catholic schools, private schools or out of New York schools were assessed on reading measures other than the MAT, SAT, or SRA, those scores are not included in the analysis. Only four or five such scores were obtained. SRA scores at the fifth grade level were considered equal to SAT scores after discussions with the respective research departments of the publishers concerned. The 15 SRA scores included with the SAT scores from the public school system are proportionately distributed between the several cells of the research design.

Two indices of reading achievement are reported in the results: the percentage of children reading at grade level or better, and the average reading score. Analyses related to the former are by Chi Square; those related to the latter by a two-tailed t-ratio. The original hypotheses of the study predicted the CT group would outperform the DT group, and the MC group would outperform the LC group, and that both would outperform the controls. The two-tailed test of significance was also used

in the analysis for age of training, for which no original hypothesis was made with respect to direction. This is recognized as a flaw in the analysis as it exists at this time.

An average reading score of 5.1 indicates that the group reading achievement is five years and one month as compared to the national norm of, in the case of the SAT fifth grade results, 5.7. Only the ten months of the school year are reflected in the scores; that is, the next higher score to 5.9 is 6.0, not 5.10.

Chi square and t-ratio analyses are detailed in Appendix B. Percent above average or better refers to the percent reading above the national average on the SAT or MAT.

A. Reading as a Function of Age of Training

Table II shows the reading level of the subjects who participated in the Harlem study, and of their Comparison Group, during the years 1975, 1974 and 1973 when the modal subject in the sample was in grades five, four and three, respectively. For each grade level or above, and for average reading score.

1. Reading and Age of Training: Grade 5 (1975)

a) Percent at grade level or better. Forty eight per cent of T-2, 40% of T-3, 36% of the Controls, and 31% of the Comparison Group were reading at grade level or better in April, 1975, when the measure was given. Chi Square analysis revealed that the proportion of children in T-2 reading at grade level or better was significantly higher than that proportion of the Comparison Group ($p < .03$). Other possible comparisons did not reach the .05 level of statistical significance. The T-2 Group at 48% is only

slightly below the national norm for the SAT (50%).

b) Average Reading Score: The average reading scores were T-2=5.35, T-3=5.02, Control=5.09 and Comparison=5.09. None of the t-ratios between groups achieved statistical significance although the T-2 vs. Comparison Group approached it. (See Appendix B.)

2. Reading and Age of Traing: Grade 4 (1974)

a) Percent at Grade Level or better. Thirty percent (30%) of T-2, 26% of T-3, 23% of the Controls, and 22% of the Comparison Group read at grade level or better. Chi Square revealed no significant differences although T-2 vs. comparison approached the .05 level of confidence (Chi Square = 3.73 df1: critical value=3.80).

b) Average reading score: Average reading scores were T-2=4.06, T-3=3.79, Controls=3.70, and Comparison=3.79 (both boys and girls). No t-ratio between pairs was statistically significant.

3. Reading and Age of Training: Grade 3 (1973)

a) Percent at Grade Level or better: forty seven percent (47%) of T-2, 40% of T-3, 26% of Controls, and 29% of Comparison were reading at grade level or better. Chi Square revealed T-2 to be significantly higher than Comparisons. No other pairings reached a satisfactory level of confidence but several approached that level.

b) Average Reading Score: Average reading scores were T-2=3.60, T-3=3.40, Control=3.39, and Comparison=3.35. None of the differences were statistically significant.

4. Discussion: Reading and Age of Training

The T-2 Group had the highest average reading score in grades

TABLE II

AGE OF TRAINING AND READING AT AGES 9-11

	1 Trained At 24 Months	2 Trained At 36 Months	3 Control Group	4 Comparison Group
1975: 5th Grade (SAT)	(N = 40)	(N = 53)	(N = 28)	(N = 852)
a) % Grade Level or Better (5.7 Years)	48%	40%	36%	30%
b) Average Reading Score	5.35	5.02	5.09	5.01
1974: 4th Grade (MAT)	(N = 40)	(N = 45)	(N = 30)	(N = 2096)
a) % Grade Level or Better	30	25	23	22
b) Average Reading Score	4.06	3.79	3.70	3.79
1973: 3rd Grade (MAT)	(N = 36)	(N = 43)	(N = 19)	(N = 2160)
a) % Grade Level or Better	47	40	26	29
b) Average Reading Score	3.60	3.40	3.39	3.35

3, 4, and 5. The probability of that occuring by chance is .0156. At grade three T-2 was 2.5 months ahead of Comparison and 2 months ahead of T-3 and Control. At grade 4, 3 months ahead of Comparison and Control and 1.5 months ahead of T-3. At grade 5 they were 3.5 months ahead of T-3, Comparison and Control.

At each grade a higher proportion of T-2 reads at grade level or better. The difference between T-2 and Comparison is statistically significant at grades 3 and 5 and misses significance at grade 4 by .07 from the critical value of 3.80.

While T-3 shows a higher proportion reading at grade level than Control or Comparison at each grade level, their average scores are almost identical to those two groups at grades 3 and 5.

B. Reading and Type of Training

Table III shows reading in the three grades as a function of whether subjects received CT or DT training. The Control and Comparison Group scores shown there are, of course, the same as shown in Table II.

1. Reading and Type of Training: Grade 5 (1975)

a) Percent above grade level: Forty seven percent (47%) of CT, 37% of DT, 36% of Controls, and 30% of Comparisons read at grade level or above. CT was significantly better than Comparison ($p = < .02$). No other comparisons between groups was significant.

b) Average reading score: The average reading score for CT was 5.29, DT was 4.97, Control 5.09, and Comparison was 5.01. None of the t-ratios was significant at the .05 level of confidence.

2. Reading and Type of Training: Grade 4 (1974)

a) Percent above average or better: Twenty nine percent (29%)

TABLE III

TYPE OF TRAINING AND READING AT AGES 9-11

	Concept	Discovery	Control	Comparison
1975: 5th Grade (SAT)	(N = 55)	(N = 36)	(N = 28)	(N = 852)
a) % Grade Level or Better	47	37	36	30
b) Average Reading Score	5.29	4.97	5.09	5.05
1974: 4th Grade (MAT)	(N = 47)	(N = 37)	(N = 28)	(N = 2096)
a) % Grade Level or Better	29	21	23	22
b) Average Reading Score	3.95	3.86	3.70	3.79
1973: 3rd Grade (MAT)	(N = 45)	(N = 29)	(N = 19)	(N = 2160)
a) % Grade Level or Better	43	43	26	29
b) Average Reading Score	3.60	3.40	3.39	3.35

of CT, 21% of DT, 23% of Control, and 22% of Comparison were reading at the national average for grade 4 or better. Chi Square analysis shows the proportion of CT's reading to that criteria is significantly higher than Comparison. No other difference between groups was statistically significant.

b) Average reading score: The average reading score for CT was 3.95; for DT it was 3.86, for Controls 3.70 and for Comparison, 3.44. No Comparisons were statistically significant.

3. Reading and Type of Training: Grade 3 (1973)

a) Percent above average or better. Forty three percent (43%) of CT, 43% of DT, 26% of Controls and 29% of Comparisons were reading at grade level or better. The difference between CT and Comparison was statistically significant at the $p \leq .05$ level of significance. Other differences were not but some approached that criterion (See Appendix B).

b) Average Reading Scores: CT (3.60) was 2.5 months ahead of Comparison (3.35), and two months ahead of DT (3.40) and Controls (3.39). No differences between groups was statistically significant.

4. Discussion: Type of Reading

The CT Group had the highest percent reading above grade level in Grade 5 and was equal to DT and higher than Control and Comparison in Grades 3 and 4. The probability of that occuring is .031. At grade three, the average number of the CT was 3.5 months ahead of Comparison, 1 month ahead of Control, and equal to DT. At grade 4, CT was 5 months ahead of Comparison, 2.5 months ahead of Control, and one month ahead of DT. At grade 5, CT was 3 months ahead of Comparison, 2 months ahead of Control, and 3 months ahead of DT.

The DT Group, better than Control or Comparison at Grade 3, and still better than Control and Comparison at grade 4 while no less than at grade three is no better at grade 5 than either Control on average reading score or percent reading at grade level or above.

C. Reading by Social Class:

Comparison between Social Class across treatment groups is shown in Tables IV and V. Table IV shows percent at grade level and average reading score, and Table V details the relationship between Social Class and treatment (CT, DT, C) and percent at grade level or better.

1. Reading and Social Class: Grade 5 (1975)

a) Percent at age level or better: Chi Square 4 x 2 analysis reveals that the four conditions have significantly different proportions of subjects reading at grade level or better ($p = .025$). Analysis of dyads shows the proportion in the MC to be significantly different from the Comparison Groups ($p = .025$). No other dyad comparisons were statistically significant.

The MC x CT x T-2 cell and the MC x CT x T-3 cell have the highest percentage reading at grade level or better of all possible cells, and the LC x DT x T-3 cell is the lowest. The LC x DT x T-2 group has 45% at that criteria, only slightly below the national norm.

b) Average reading score: The t-ratios (Appendix B) show that MC has significantly higher reading scores than LC ($t=2.20$, $p = .05$) and the Comparison Group ($t=2.24$, $p = .05$). No other comparisons were significantly different.

2. Reading and Social Class: Grade 4 (1974)

a) Percent at Grade Level or better: A Chi Square 4 x 2 analysis

TABLE IV

SOCIAL CLASS AND READING AT AGES 9-11

	Lower Class	Middle Class	Control	Comparison
1975: Fifth Grade (SAT)	(N = 63)	(N = 58)	(N = 27)	(N = 852)
a) Grade Level or Better	35	48	36	30
b) Average Reading Score	4.80	5.53	5.09	5.01
1974: 4th Grade (MAT)	(N = 59)	(N = 60)	(N = 20)	
a) Grade Level or Better	20	37	25	
b) Average Reading Score	3.62	4.10	3.70	3.79
1973: 3rd Grade (MAT)	(N = 41)	(N = 38)	(N = 19)	
a) % Grade Level or Better			26	
b) Average Reading Score	3.49	3.50	3.39	3.35

TABLE V

PERCENT READING AT GRADE LEVEL, GRADE 5:
AGE LEVEL X TYPE OF TRAINING X SOCIAL CLASS

	Concept Training	Discovery	Control
A. Trained at Two:			
Lower Class	45% (N = 11)	50% (N = 8)	33% (N = 12)
Middle Class	42 (N = 12)	67 (N = 6)	40 (N = 15)
B. Trained at Three:			
Lower Class	36 (N = 14)	19 (N = 16)	33
Middle Class	63 (N = 14)	50 (N = 6)	40
C. A and B			
Lower Class	40 (N = 25)	29 (N = 24)	33
Middle Class	54 (N = 26)	57 (N = 12)	40

reveals the proportions at grade level or better for the four conditions differs significantly ($p \leq .05$). No differences existed between dyads.

b) Average reading score: Appendix B shows that while no dyad comparisons yielded statistically significant differences, several approached the critical value. MC had the highest average reading score, 4.11, 7 months ahead of the Comparison Group, 4 months ahead of Control, and 3.5 months ahead of LC. LC bested the Comparison Group by two months.

3. Reading and Social Class: Grade 3 (1973)

a) Percent above grade level: The proportions reading above grade level did not differ significantly across or within the four groups in the analysis. The LC (40%) was equal to the MC (40%) in subjects reaching the grade level criterion.

b) Average reading score: No differences exist by Social Class for the average reading scores in the groups analyzed.

4. Discussion: Reading and Social Class

The LC sample appears to have benefitted from their involvement in the study despite the lack of statistical significant comparisons. Their average reading score is 4 months better than the Comparison Group in the third grade (comprised of boys and girls, both MC and LC), two months ahead of the average for a similar Comparison Group in the fourth grade and lage behind the Comparison Group in the 5th grade - but has a higher percentage reading at grade level or better. They are 2 months behind the Control Group in Grade 5 (both MC and LC) but are equal to that Group in percent reading at grade level. At grade 4 they are only slightly below the Controls, for both indices; and at Grade 3, they read on both indices.

Not surprisingly, the superiority of the MC group appears to increase with age.

D. Discussion: Reading

Because of the lack of statistical significance of some comparisons between T-2 and CT on the one hand and Control and Comparison on the other, we are not absolutely certain that those two conditions have significantly affected the reading scores of the children who comprise them. More fifth grade scores will be obtained in the next follow-up (1977) and with the power of additional subjects, perhaps those questions will be resolved.

All of the data, however, points to three conclusions about reading: (1) Training at age two is more effective than training at age three or no training at all. The T-2 Group is reading at grade level only slightly below the national norms (48% vs. 50%). Its average reading score for grade 5 is 3.5 months below that norm of 5.7, but is 3 months ahead of Controls, Comparisons, and even T-3. The probability that chance would explain their superiority at each grade level is 15 to 1,000. At each grade level, T-2 exceeds all other groups in percent reading at grade level. These data cannot speak to why intervention at age two is more effective than intervention at age three for reading level in grade 5.

(2) Concept Training is more effective than no training or discovery training. The CT group is reading at grade 5 only slightly below the national norms. Its average reading score is the highest at every grade. At grade 5, the average reading score is 5.3, which looks poor compared to the national norm, but good when compared to the Control (5.1) and Comparison Groups (5.0). For the boys in that Group, they have almost caught up with the average for their female peers (5.3 vs. 5.4).

Why those in the DT Group have lost ground with respect to reading, when at 4/8 they were equal to the CT group on almost every measure , is fascinating to consider - but these data contribute little to why.

(3) The Lower Class children in the training groups seem to be holding their own when compared to their Control peers to include MC and girls. This statement is perhaps more speculative than (1) and (2) above. But their percent reading at grade level and average scores compare favorably with Controls and Comparisons at every level except the fifth grade average score, and even there, they are equal to Controls and better than Comparisons on percent reading at grade level.

Beyond those conclusions, there are some significant aspects to the data.

The small (N=28) sample of Control subjects found appear to be representative of the Comparison Group (N=852). Percent reading at grade 5 level (36% vs. 30%) and average reading score (5.09 to 5.01) are only slightly higher in the Controls than in the Comparisons. Considering the Hawthorne effect and their experience in previous assessments, one might have predicted a greater difference. But, perhaps Hawthorne effects and experience with psychological tests are unrelated to whatever reading is. In any event, the Controls appear representative of the Harlem population - so that generalizations from differences found between T-2 and CT and Comparisons to differences between those groups and the Controls may not be as presumptuous as they might first appear.

Finally, a note about the distributions of reading scores - those in the study are highly variable.

But the standard deviation of 5th grade reading scores in our sample is highly related to the size of the N in each condition:

GROUP	N	SD
Comparison	852	1.51
T-3	53	1.61
L-C	63	1.74
CT	55	1.78
T-2	40	1.78
DT	38	1.83
MC	58	1.92
Control	28	2.06

for a Spearman Rank order correlation of .75. Thus, we may expect that as the N is increased in the sample in subsequent follow-ups, the SD of each group will diminish as the N gets larger. When the mean differences in the t-ratio remain the same, the N gets larger, and the SD gets smaller- the ratio gets larger and is more likely to be at a satisfactory level of significance. Thus, if the mean differences remain the same, we may expect more statistically significant differences between groups with subsequent reading scores - because with more scores, the N gets larger and the SD smaller. Furthermore, as the sample grows with subsequent follow-up, cells of the original design (e.g. LC x T-2 x CT) will become adequate in size for multivariate analysis, more powerful statistics for determining the main effects of this study as well as interactions within those effects. Hopefully, we shall be able to make such statements as "for lower class children to read better the data indicate that condition (T-2) + (CT) is

best, but for MC children either (T-2)+ (CT) or (T-3) + (CT) work equally well."

IV. IQ RESULTS

The intelligence test scores (IQ) scores reported below are derived from the Wechsler Intelligence Scale for Children (WISC). They are combined scores taken from the administration of two revisions of that measure, the WISC (1949) and the WISC-R (1974).

Ninety (90) children of the original sample were administered the WISC in 1974, when they were ten years of age. Forty-nine (49) were administered the WISC-R in 1976, when they were twelve years of age.

The scores have been combined using corrections from the WISC-R to the old WISC. Those corrections were obtained from the research section of the Psychological Corporation. ( ) They are: For the Full Scale WISC IQ, 2.5 points; for the Verbal IQ, zero points; and for the Performance IQ, 4.6points. The corrections are made by adding 2.5 or 4.0 to the Full and Performance IQ's obtained in the 1976 assessment. The norms with which the scores listed below are correctly compared are the norms on the old WISC.

One reason the old WISC was revised was because the norms originally derived, a mean IQ of 100, changed to about 105. The revision presumably means the average score on the WISC-R will more closely approximate 100. Our transformation to old WISC scores was made to conform more nearly to the norm the educational and psychological communities as well as the public use implicitly when hearing average IQ scores.

Poor, Black children in the public schools have characteristically averaged around 91 on the old WISC norms, , although that average changes with age downward. Thus, we may expect that the average score for samples of such children will be lower on the WISC-R. Concerned investigators and the public at large should be informed of this change. The norms have been changed. The children are not dumber.

Why did we choose WISC-R to administer in 1976 when we had used the WISC in 1974? There are two answers. (1) In 1975, a consortium of investigators of early childhood intervention agreed to adopt some common measures, one of which is the WISC-R. Those common measures will provide the consortium, the supervisory center for which is at Cornell University under the direction of Professor Irving Lazar, with the ability to compare results across studies. The Harlem study is a member of that consortium. (2) The second reason for adopting the WISC-R is funds are available for reassessing the Harlem sample found and locating additional subjects. In 1976-77 the 90 subjects who comprised the 1974 assessment will be evaluated again, with the WISC-R. At that time, the entire sample will have scores obtained from the same measure, and corrections will not be necessary.

The 90 subjects administered the WISC in 1974 were assessed at the New York Medical College by Dr. Miriam John and her staff. She nor her staff knew which children belonged to the several cells in the research design, or which were treatment children and which were controls.

The 49 subjects administered the WISC-R in 1976 were assessed at what was once the Harlem Training Center, now the CUNY Center for Community Research and Services. The examiners (2) were in the employ of the writer, but had no knowledge of what subjects were in what cell of the design -

or, for that matter, about the design itself.

The WISC data are presented without a Comparison Group because the WISC-R is relatively new, and adequate norms on a national sample are not yet available, particularly for Black boys. Presumably, subsequent follow-up study have norms available which will provide broader comparisons.

The IQ data is presented as average scores are related to age of training, type of training and social class. Statistical analyses are limited to the t-ratio as described in the analysis of the reading scores.

Statistical analyses (t-ratio) for groups compared by average IQ score are in Appendix F.

A. IQ and Age of Training

Table VI presents Full Scale, Verbal and Performance IQ's from the WISC. The former is derived by combining the scores obtained on verbal and performance.

1. Full Scale IQ: No significant difference exists between the average IQ of T-2 ($\bar{x}$=99.40) and T-3 ($\bar{x}$=99.36). The difference between T-3 and Control ($\bar{x}$=99.36 vs. 93.16) is statistically significant at the $p \leq .05$ level (t=2.00, df=84). No statistical difference exists at a satisfactory level of confidence (.05) between T-2 and Control, but that difference approaches that value ($p \leq .10$; t=1.80, df=78; critical value = 2.00 (See A. 3., Performance).

2. Verbal IQ: No significant differences exist between T-2 (100.32), T-3 (99.97) and Control (96.26).

3. Performance IQ: Both T-2 (t=2.15, df=80) and T-3 (t=2.78, df=86) are significantly better than the Controls. No difference exists between

Table IV

IQ and Age of Training

	Trained at 24 mos. (N=53)	Trained at 36 mos. (N= 59)	Control group (N= 27)
WISC-full			
$\bar{X}$	99.40	99.36	93.16
SD	13.79	11.78	16.27
Verbal			
$\bar{X}$	100.32	99.97	96.26
SD	15.78	13.99	17.89
Performance			
$\bar{X}$	97.92	98.82	90.76
SD	14.00	11.68	14.10

T-2 and T-3. Since no differences were found on the Verbal Scale between T-2 and T-3, and Control - we conclude that the Full Scale IQ differences found (A.1) are largely a function of those items of the WISC which contribute to the Performance Score.

B. IQ and Type of Training

Table VII presents means and standard deviations for IQ and type of training.

1. Full Scale IQ: No differences exist between CT (100.16) and DT (98.34). The difference between CT (100.16) and Control (93.16) is significant at the $p \leq .025$ level of confidence ($t=2.24$, $df=89$). The difference between DT and Control is not statistically significant ($t=1.50$, $df=73$).

2. Verbal: No significant differences exist between CT (100), DT (100) and Control (96) on the verbal measure despite the four point score difference.

3. Performance: CT was significantly different from Control ($p < .005$). DT was significantly different from Control ($p < .05$). No difference existed between CT and DT. Both types of training influenced the Performance IQ significantly.

C. IQ and Social Class

Table VIII presents IQ data as a function of Social Class.

1. Full Scale: MC differs significantly from LC ($t=2.47$, $df=84$); MC differs significantly from Controls ($t=2.74$, $df=73$).

2. Verbal: MC (102) is significantly higher than LC (98), $p \leq .05$. LC nor MC differ significantly from the Controls.

Table VII

IQ and Type of Training

	Concept Training (N=64)	Discovery Training (N=48)	Control Group (N=27)
Full Scale			
$\bar{X}$	100.16	98.34	93.16
SD	12.34	13.25	16.27
Verbal			
$\bar{X}$	100.36	99.83	96.26
SD	14.40	15.38	17.89
Performance			
$\bar{X}$	99.38	97.07	90.76
SD	12.22	13.51	14.10

Table VIII

IQ and Social Class

	Participating		
	Lower (N=53)	Middle (N=59)	Control (N=27)
Full Scale			
$\bar{X}$	96.31	102.14	93.16
SD	11.73	13.02	16.27
Verbal			
$\bar{X}$	97.62	102.39	96.26
SD	13.99	15.19	17.89
Performance			
$\bar{X}$	95.79	101.08	90.76
SD	12.18	12.48	14.10

3. Performance: MC (101) is significantly higher than LC (96), $p \leq .025$. MC is significantly higher than Controls (91), $p \leq .005$. LC (96) approaches significance at the .05 level when compared to the Controls ($t=1.66$).

IQ: DISCUSSION

Both ages of training (T-2 or T-3) affected full IQ significantly. When compared to the Control Group, Concept Training but not Discovery Training affected IQ significantly as well. But the data are clear with respect to the subscale of the WISC which contributed to that IQ difference. There were no significant differences between groups for age or type of training on the Verbal Scale, but the differences on the Performance Scale are consistent and impressive. The respective IQ's reflecting that scale were T-2=98, T-3=99, CT=99 and DT=97, all of which are significantly higher than the Control Performance IQ of 91. The domain of behavior which the items on the Performance Scale measure is the domain influenced by our early intervention.

There were no significant differences at age 4/8 in favor of type or age of training as compared to Control. The fact that the sample size was almost twice as large at 4/8 makes the argument for a "sleeper effect" persuasive - on the same test (WISC) given at 4/8 and at ages 10 and 12 - no differences occurred at 4/8 but were found later.

While the IQ evidence shows no differences by age of training at this time - there is still some suggestion that T-2 is more effective than T-3. The T-2 Stanford Binet IQ average at 2/8, <u>after training</u>, was

93. A year later, at 3/8, it was 97. The T-3 average at 3/0, before training, was 96. At 3/8, after training, it was 99. Thus, T-2 changed more. T-3 began with higher IQ's, but T-2 is higher now - but not significantly so.

The original IQ advantage at 3/0 for T-2 is also relevant in interpreting results related to type of training. The T-2 group was comprised of half DT's and half CT's, as was the T-3 Group. Thus, even if T-3 had continued to excede T-2 at age 12, the early IQ advantages of T-3 would not influence the type of treatment results.

Both types of training influenced IQ. The major influence was on whatever it is that the Performance Scale measures, rather than the Verbal Scale. There are nine IQ points difference on the Performance Scale between Concept Training and Control!

V. General Discussion and Conclusions

The data are persuasive that at least one form of intervention, Concept Training, at age 24 or 36 months significantly affects reading the the fifth grade and IQ at age 10-12. Intervention at age two had an effect on reading and IQ - whereas intervention at age three affected IQ, but not reading. It is also clear that Discovery Training affected IQ, but did not affect reading.

The evidence on the IQ is conclusive and illuminating. The IQ's were affected because of what the Performance Scale measures, which is different from what the Verbal Scale measures it cannot be said that the data are statistically significant "but not really significant". Nine IQ points on the Performance Scale, 7 IQ points on the Full Scale for the

64 CT children is significant, in this case, regardless of the connotation of the word significant, that for an introduction two hours weekly for eight months, ten years ago.

The evidence for the effect on reading is conclusive only if one is persuaded by two arguments: (1) that the Comparison Group is a valid Control Group - equal in Social Class and intellective potential to the Controls involved in the study and without the possible advantages of prior testing and the Hawthorne effect and (2) that the combined effect of many t-ratios at or approaching statistical significance for the T-2 and CT groups and the probability of those groups leading all other groups in average reading score at all three grade levels ($p = .015$) provide a level of statistical confidence which no single t-ratio does provide.

The Comparison Group included all boys in the fifth grade in District 5 (Harlem) whose names on class rosters left no doubt as to gender. The original Control Group was comprised of 55% Category V; 34% Category IV; 7% Category III; 4% Categories II and I on the Hollingshead-Redlich. Harlem, despite its reputation, is not all ghetto - many Black middle class families attend school there. Presently, the unemployment rate in Harlem is close to 30%. A larger proportion of our original sample were unemployed when the study began. The attrition analysis shows that the

sample found and assessed in 1976 did not differ by Social Class from those in the original sample who were not found and assessed (Appendix A,II). If anything, it appears that the sample in this study was somewhat lower in SES than the Harlem population of fifth grade boys of which it is a subset.

The cumulative effect of repeated measures each of which is just short of the .05 level of confidence is best demonstrated in Appendix B, I, B, 2a. The 4 x 2 Chi Square is 7.08, 6.26, and 6.55 for grades 5, 4, and 3, respectively. The critical value for .05 with three degrees of freedom is 7.8. All three are at a p= .10 level of confidence. It is argued that those proportions do differ across groups at a satisfactory level of statistic confidence. If the Concept Training Group had not had the highest average score at each grade, we might conclude that the distributions differ but be uncertain as to what group read best. But the chance that CT would read best at every age given four possible winners, is .015.

The t-ratios between T-2 and CT on the one hand and Control and Comparison Groups are consistently in favor of the former groups on every comparison. For those to reach a satisfactory level of confidence, three events must occur as additional subjects are found and assessed: (1) The sample size must increase. (2) The standard deviations must decrease, and (3) the means must remain the same or increase.

Additional subjects assessed is synonomous with the sample size increasing. The SD will almost certainly decrease as the N increases: (a) within groups the rho between sample sizes and SD's is -.75; (b) the SD for the 852 boy Comparison Group is 1.51, for the Controls it is 2.02; CT is 1.78; and T-2 is 2.02. The sample in the study is a subset of the

Comparison Group. The only possibility which can prevent increased statistical confidence with additional subjects is that the difference between means decreases - that, of course, is what the t-ratio states. Presently, our distributions are so variable that the mean for each group may change with additional subjects, in either direction.

If one is not persuaded by these arguments that the combined effects of the reading analysis are conclusive about the superiority of the T-2 and CT groups - what can one conclude? Certainly not that the intervention has failed. Only that more subjects should be assessed before such conclusiveness is shown, and we are continuing to locate and assess subjects presently not found.

For those who are persuaded the evidence is conclusive and those who grant that a good case has been made but have reservations about the data being conclusive, there are some important questions raised by the data.

Why is Concept Training so much more effective than Discovery Training for subsequent reading, when the latter has an equal effect on the IQ? Indeed, at grades 5 and 4, the DT group is no better than Control or Comparison on that skill. It would be extremely difficult to argue the DT had any effect on reading. And yet, the DT treatment is the most similar to what occurs in most nursery schools and day care centers today. While these data refer only to treatments characterized by the one-to-one situation, it does not seem to be too great a generalization to apply the results to Group Training as well. The suggestion is that most preschool training today may well have an effect on IQ scores, but none on reading.

The suggestion is that early intervention programs should include one-to-one interaction between teacher and child - and that the training should be structured, planned, and teacher-guided.

Why is reading more effective by training at age two than training at three, where IQ scores for T-2 and T-3 are almost identical? Clearly, these results tend to support those who have argued for intervention to occur as early as possible, and raise questions about the position that intervention at any age is equally effective. But one can only speculate about why. And, this report is not the place for that.

Why were there no differences by age or type of training at age 4/8 when training at two and Concept Training yields increased IQ's and reading scores at age 12? The existence of sleeper effects must be retained as one possibility. The reading results, it may be argued, are there at 12 and not at 4/8 because no measure was given at 4/8 valid for whatever reading is. But the IQ's are different at 12 and were not at 4/8 - the WPSSI was used at 4/8 and the WISC at 12. Those measures gained part of their utility because the Verbal and Performance scales are comprised of the same kinds of items, administered in the same manner, and conceived by the same expert in test construction.

Why is the Performance IQ and not the Verbal IQ affected by the treatments? Only conjecture is possible at this time, so it will not be discussed. But the question deserves serious attention by those concerned with early intervention.

How is it possible that two hours weekly of Concept Training at age two or three for a period of eight months, with the average number of

hours in training being only 45, affected Performance IQ nine points and average reading score three months? More speculation, but it did.

And, finally, there is that word significance. At least two connotations of the word are relevant. Statistical significance, which has been discussed; and significance for the children concerned. Is it significant that 63 found children have nine points higher Performance IQ's and read three months ahead - and that we may infer that 60 others not found do so as well? We think it is.

VI. RECOMMENDATIONS

On the basis of the evidence presented here, it is recommended that early childhood compensatory education programs:

(1) Begin at an earlier age than is presently so, and,

(2) That all programs have imbedded in them periods during each week when every child receives structured, planned, teacher-guided and concept oriented one-to-one instruction.

REFERENCES

Doppelt, J. E. and Kaufman, A. A. Estimation of the Differences Between WISC-R and WISC IQ's. Journal of Educational and Psychological Measurement, In Press.

Hollingshead, A. B. The Two Factor Index of Social Position, New Haven, Conn., Author, 1957

Palmer, F. H. Socioeconomic Status and Intellective Performance Among Negro Preschool Boys. Developmental Psychology, 1970, 3, 1-9.

Palmer, F. H., Minimal Intervention at Age Two and Three and Subsequent Intellective Changes. In R. K. Parker, The Preschool in Action, Boston: Allyn and Bacon, 1972

APPENDIX A

APPENDIX A

I. Attrition analysis by 3.8 Stanford-Binet IQ[a]

	1974 - 1976 Found $\bar{X}_1$	Not Found $\bar{X}_2$	t	d.f
T-2	97.2	96.3	.32	100
T-3	101.3	97.0	1.72	108
Cont.	91.3	95.3	-1.13	58
CT	99.4	95.8	1.47	107
DT	98.0	98.1	- .02	102
L	97.5	95.5	.81	120
M	101.0	100.2	.28	93

II. Attrition analysis by Hollingshead-Redlich SES Score[a]

	1974 - 1976 Found $\bar{X}_1$	Not Found $\bar{X}_2$	t	d.f
T-2	58.9	57.6	.48	113
T-3	56.9	57.1	- .11	120
Cont.	58.0	58.3	- .11	66
CT	57.4	58.4	- .40	118
DT	59.5	56.7	-1.14	118

a. None of the above comparisons were significant at $\rho = .05$, two-tailed.

III. Attrition by Hollingshead SES Categories[a]

(# in 1974-76 sample/ # in 1966-7 sample)

	I	II	III	IV	V
Scores	11 - 17	18 - 27	28 - 48	44 - 60	61 - 77
T-2	0/1	1/6	1/5	23/42	28/69
T-3	0/0	0/3	6/13	29/42	31/66
Cont.	0/1	2/3	2/5	13/23	16/36

a. Scores and categories range from V (61-77) as the lowest SES to I (11-17) as the highest.

APPENDIX B

Appendix B: Analysis of 1975 Assessment

I. Reading

A. Reading as a function of age at training

1. t-ratios on individual reading scores

a. Grade 5 (SAT)

	$\bar{X}_1$	$\bar{X}_2$	t	d.f.
T-2 - T-3	5.35	5.02	.87	91
T-2 - Cont.	5.35	5.09	.52	67
T-2 - Comp.	5.35	5.01	1.23	890
T-3 - Cont.	5.02	5.09	- .17	80
T-3 - Comp.	5.02	5.01	.04	903
Cont. - Comp.	5.09	5.01	.24	879

b. Grade 4 (MAT)

T-2 - T-3	4.06	3.79	.88	87
T-2 - Cont.	4.06	3.70	1.00	68
T-2 - Comp.	4.06	3.79	1.31	2134
T-3 - Cont.	3.79	3.70	.27	77
T-3 - Comp.	3.79	3.79	.00	2143
Cont. - Comp.	3.70	3.79	- .38	2124

c. Grade 3 (MAT)

T-2 - T-3	3.60	3.40	.79	77
T-2 - Cont.	3.60	3.39	.66	53
T-2 - Comp.	3.60	3.35	1.24	2194
T-3 Cont.	3.40	3.39	.03	60
T-3 - Comp.	3.40	3.35	.27	2201
Comp. - Cont.	3.39	3.35	.14	2177

2. Chi-square statistics on differences in proportion of children reading at or above the national norm.

a. 2 x 4 Chi-Square Statistics

i. Fifth Grade:

	T-2	T-3	Cont.	Comp.	
Above Norm	19	21	10	262	$X^2 = 6.61$
Below Norm	21	32	18	590	n. s.

ii. Fourth Grade:

	T-2	T-3	Cont.	Comp.	
Above Norm	12	13	7	379	$X^2 = 6.29$
Below Norm	28	36	23	1717	n. s.

iii. Third Grade:

	T-2	T-3	Cont.	Comp.	
Above Norm	17	17	5	642	$X^2 = 7.10$
Below Norm	19	26	14	1518	n. s.

b. 2 x 2 Chi-square statistics

i. Grade 5:

	T-2	T-3	Cont.	Comp.
T-2		0.58	0.94	4.97*
T-3			0.12	1.83
Cont.				0.31
Comp.				

ii. Grade 4:

	T-2	T-3	Cont.	Comp.
T-2		0.13	0.39	3.73
T-3			0.10	2.29
Cont.				.55
Comp.				

iii. Grade 3:

	T-2	T-3	Cont.	Comp.
T-2		0.4[illegible]	2.26	5.16**
T-3			1.01	1.94
Cont.				0.10
Comp.				

* $p < .05$

** $p < .03$

B. Reading as a function of type of training

1. t-tests on individual reading scores

a. Grade 5 (SAT)

	$\bar{X}_1$	$\bar{X}_2$	t	d.f.
CT - DT	5.29	4.97	.84	91
CT - Cont.	5.29	5.09	.46	82
CT - Comp.	5.29	5.01	1.18	905
DT - Cont.	4.97	5.09	- .25	65
DT - Comp.	4.97	5.01	- .14	888
Cont. - Comp.	5.09	5.01	+ .24	879

b. Grade 4 (MAT)

CT - DT	3.95	3.86	.29	87
CT - Cont.	3.95	3.70	.75	79
CT - Comp.	3.95	3.79	.87	2145
DT - Cont.	3.86	3.70	.43	66
DT - Comp.	3.86	3.79	.33	2132
Comp. - Cont.	3.70	3.79	- .38	2124

c. Grade 3 (MAT)

CT - DT	3.49	3.50	- .04	77
CT - Cont.	3.49	3.39	.32	66
CT - Comp.	3.49	3.35	.81	2207
DT - Cont.	3.50	3.39	.33	47
DT - Comp.	3.50	3.35	.68	2188
Cont. - Comp.	3.39	3.35	.14	2177

2. Chi-square statistics on differences in proportion of children reading at or above the national norm.

a. 2 x 4 Chi-Square Statistics

i. Fifth Grade:

	DT	CT	Cont.	Comp.	
Above Norm	14	26	10	262	$X^2 = 7.08$
Below Norm	24	29	18	590	n. s.

ii. Fourth Grade:

	DT	CT	Cont.	Comp.	
Above Norm	10	15	7	379	$X^2 = 6.26$
Below Norm	28	36	23	1717	n. s.

iii. Third Grade:

	DT	CT	Cont.	Comp.	
Above Norm	13	21	5	642	$X^2 = 6.55$
Below Norm	17	28	14	1518	n. s.

b. 2 x 2 Chi-square statistics

i. Grade 5:

	CT	DT	Cont.	Comp.
CT		1.00	1.01	6.51**
DT			0.01	0.63
Cont.				0.31
Comp.				

ii. Grade 4:

	CT	DT	Cont.	Comp.
CT		.10	.35	4.27*
DT			.08	1.70
Cont.				.55
Comp.				

iii. Grade 3:

	CT	DT	Cont.	Comp.
CT		0.00	1.59	3.94*
DT			1.45	2.61
Cont.				0.10
Comp.				

* $p < .05$

** $p < .025$

C. Reading as a function of SES

1. t-tests on individual reading scores

a. Grade 5 (MAT)

	$\bar{X}_1$	$\bar{X}_2$	t	d.f.
L - M	4.85	5.54	-1.86*a	91
L - Cont.	4.85	5.09	- .57	78
L - Comp.	4.85	5.01	- .66	901
M - Cont.	5.54	5.09	.95	69
M - Comp.	5.54	5.01	1.97**a	892
Cont. - Comp.	5.09	5.01	.24	879

b. Grade 4 (MAT)

L - M	3.70	4.12	-1.37	87
L - Cont.	3.70	3.70	.00	72
L - Comp.	3.70	3.79	- .46	2138
M - Cont.	4.12	3.70	1.26	73
M - Comp.	4.12	3.79	1.70	2139
Cont. - Comp.	3.70	3.79	- .38	2124

c. Grade 3 (MAT)

L - M	3.52	3.46	.25	77
L - Cont.	3.52	3.39	.40	58
L - Comp.	3.52	3.35	.90	2199
M - Cont.	3.46	3.39	.23	55
M - Comp.	3.46	3.35	.56	2196
Cont. - Comp.	3.39	3.35	.14	2177

* $P < .05$
** $P < .025$
a One-tailed

2. Chi-square statistic for differences in proportion of children reading at or above the national norm.

a. 2 x 4 Chi-Square Statistics

i. Fifth Grade:

	L	M	Cont.	Comp.
Above Norm	18	22	10	262
Below Norm	33	20	19	590

$X^2 = 8.96$
$P < .05$

ii. Fourth Grade:

	L	M	Cont.	Comp.
Above Norm	9	16	7	379
Below Norm	35	29	23	1717

$X^2 = 9.47$
$P < .025$

iii. Third Grade:

	L	M	Cont.	Comp.
Above Norm	18	16	5	642
Below Norm	23	22	14	1518

$X^2 = 6.58$
n. s.

b. 2 x 2 Chi-square statistics

i. Grade 5:

	L	M	Cont.	Comp.
L		2.74	0.01	0.46
M			2.22	8.64**
Cont.				0.18
Comp.				

ii. Grade 4:

	L	M	Cont.	Comp.
L		2.51	.09	.16
M			1.26	8.94**
Cont.				0.55
Comp.				

iii. Grade 3:

	L	M	Cont.	Comp.
L		.03	1.70	3.85*
M			1.36	2.73
Cont.				0.10
Comp.				

* $p < .05$

** $p < .01$

II. IQ

A. WISC full scale scores

1. t-tests between treatment groups

	$\bar{X}_1$	$\bar{X}_2$	t	d.f.
T-2 - T-3	99.40	99.36	.02	110
T-2 - Cont.	99.40	93.16	1.80*	78
T-3 - Cont.	99.36	93.16	2.00**	84
CT - DT	100.16	98.34	.75	110
CT - Cont.	100.16	93.16	2.24++	89
DT - Cont.	98.34	93.16	1.50	73
L - M	96.31	102.14	-2.47++	111
L - Cont.	96.31	93.16	.99	78
M - Cont.	102.14	93.16	2.74++	84

* $p < .05$, One-tailed

** $p < .025$, One-tailed

++ $p < .01$, One-tailed

B. WISC verbal scale score

1. t-tests between treatment groups

	$\bar{X}_1$	$\bar{X}_2$	t	d.f.
T-2 - T-3	100.32	99.97	.13	110
T-2 - Cont.	100.32	99.26	1.04	78
T-3 - Cont.	99.97	96.26	1.04	84
CT - DT	100.36	99.83	.19	110
CT - Cont.	100.36	96.26	1.53	89
DT - Cont.	99.83	96.26	.98	73
L - M	97.62	102.39	-1.72*	110
L - Cont.	97.62	96.26	.37	78
M- Cont.	102.39	96.26	1.64	84

* $p < .05$, One-tailed

C. WISC performance scale score

1. t-tests between treatment groups

	$\bar{X}_1$	$\bar{X}_2$	t	d.f.
T-2 - T-3	97.92	98.82	.37	110
T-2 - Cont.	97.92	90.76	2.16**	78
T-3 - Cont.	98.82	90.76	2.78++	84
CT - DT	99.38	97.07	.95	110
CT - Cont.	99.38	90.76	2.94++	89
DT - Cont.	97.07	90.76	1.91*	73
L - M	95.79	101.08	-2.27**	110
L - Cont.	95.79	90.76	1.66	78
M - Cont.	101.08	90.76	3.42++	84

* $p < .05$

** $p < .025$

++ $p < .005$

III. Descriptive statistics

A. Age 10-12 (WISC)

	N	$\bar{X}$	S. D.
CT	64	100.16	12.34
DT	48	98.34	13.25
T-2	53	99.40	13.79
T-3	59	99.36	11.78
L	53	96.31	11.73
M	59	102.14	13.02
Cont.	27	93.16	16.27

B. Grade 5 (SAT)

CT	55	5.29	1.78
DT	38	4.97	1.83
T-2	40	5.35	2.02
T-3	53	5.02	1.61
L	51	4.85	1.66
M	42	5.54	1.91
Cont.	29	5.09	2.06
Comp.	852	5.01	1.69

C. Grade 4

	N	$\bar{X}$	S. D.
CT	51	3.95	1.40
DT	38	3.86	1.50
T-2	43	4.06	1.46
T-3	49	3.79	1.41
L	44	3.70	1.52
M	45	4.12	1.33
Cont.	30	3.70	1.53
Comp.	2096	3.79	1.29

D. Grade 3

CT	49	3.49	1.14
DT	30	3.50	1.11
T-2	36	3.60	1.09
T-3	43	3.40	1.15
L	41	3.52	1.17
M	38	3.46	1.08
Cont.	19	3.39	1.16
Comp.	2160	3.35	1.20