R E P O R T R E S U M E S

ED 020 794 PS 001 003

AN EVALUATION AND FOLLOW-UP STUDY OF SUMMER 1966 HEAD START CHILDREN IN WASHINGTON, D.C.

BY- CLINE, MARVIN DICKEY, MARGUERITE
HOWARD UNIV., WASHINGTON, D.C.

REPORT NUMBER OEO-1490 PUB DATE 68
EDRS PRICE MF-$0.50 HC-$3.32 81P.

DESCRIPTORS- *CULTURALLY DISADVANTAGED, *DISADVANTAGED YOUTH, COGNITIVE ABILITY, LOW ACHIEVEMENT FACTORS, *ACADEMIC ABILITY, *PERFORMANCE FACTORS, COGNITIVE DEVELOPMENT, INTELLIGENCE TESTS, INTELLIGENCE DIFFERENCES, INTELLIGENCE QUOTIENT, INTELLIGENCE FACTORS, BEHAVIOR RATING SCALES, PRESCHOOL LEARNING, PRESCHOOL PROGRAMS, *KINDERGARTEN CHILDREN, PRESCHOOL CHILDREN, COMPARATIVE ANALYSIS, CONTROL GROUPS, HEAD START, DISTRIC OF COLUMBIA, STANFORD BINET, PPVT, BEHAVIOR INVENTORY, METROPOLITAN READING READINESS TEST,

THE EXPERIMENTAL GROUP IN THIS STUDY WAS 112 KINDERGARTEN CHILDREN FROM 11 HEAD START CENTERS. IN ORDER TO ASSESS THE VALUE OF THE HEAD START PROGRAM, THE MEASUREMENT OF THE EXPERIMENTAL GROUP TAKEN DURING THE FALL WAS COMPARED TO A MEASUREMENT OF NON-HEAD START KINDERGARTEN CHILDREN TESTED ABOUT THE SAME TIME. AT LEAST FOUR MONTHS AFTER THE PRETESTING, POSTTESTS WERE ADMINISTERED. THE PRETEST BATTERY CONSISTED OF THE STANFORD-BINET (S-B) AND TWO FORMS OF THE PEABODY PICTURE VOCABULARY TEST (PPVT). THE POSTTEST BATTERY CONSISTED OF (1) S-B, (2) PPVT, BOTH FORMS, (3) THE BEHAVIOR INVENTORY, AND (4) THE METROPOLITAN READING READINESS TEST. THE MOST SIGNIFICANT RESULTS WERE (1) NO DIFFERENCES WERE FOUND BETWEEN THE SCORES OF THE EXPERIMENTAL GROUP AND THE SCORES OF THE CONTROL GROUP, (2) THE CHILDREN SHOWED SIGNIFICANT IMPROVEMENT BETWEEN PRETEST AND POSTTEST PERIODS ONLY ON THE RECEPTIVE FORM OF THE PPVT, (3) THE CHILDREN CONSISTENTLY DEMONSTRATED HIGHER SCORES ON THE S-B THAN ON THE RECEPTIVE PPVT, AND (4) THE PERFORMANCE OF THE KINDERGARTEN CHILDREN APPEARED TO BE AFFECTED BY WHICH SCHOOL THEY ATTENDED AND WHICH TEACHER PRESIDED IN THE CLASSROOM. ALSO 160 HEAD START CHILDREN, WHO WERE TOO YOUNG TO START KINDERGARTEN, WERE TESTED IN THEIR HOMES AND WILL BE USED FOR A FOLLOWUP STUDY. (WD)

AN EVALUATION AND FOLLOW-UP STUDY OF SUMMER 1966 HEAD START CHILDREN IN WASH., D.C.

OEO Contract No. 1490

M.G. Cline, Principal Investigator
Marguerite Dickey

AN EVALUATION AND FOLLOW-UP STUDY OF SUMMER 1966 HEAD START CHILDREN IN WASHINGTON, D. C.

Office of Economic Opportunity Contract No.1490

Marvin G. Cline, Principal Investigator
Marguerite Dickey

Howard University
Washington, D. C.

1968

ACKNOWLEDGMENTS

The cooperation and support of a large number of individuals were involved in this study. We are particularly indebted to Dr. Dorothy Johnson, Assistant Superintendent for Elementary Schools and Mrs. Veryl Martin, Coordinator of Special Programs, of the District of Columbia Public School System, for their aid in gathering the data. Without their help, this project could not have been completed.

The preparation of the data and the statistical analyses were carried out by Gladys Malachi, Sunday Udom, Patricia Crecca, Mark Siegel and Gregory Mestanos. The supervision and training of the testers was the responsibility of Rae Ingebretson and Marilyn Cohen, both of whom also administered a number of tests. Muriel Kendrix dealt with the many organizational matters faced by projects of this sort, and she also was responsible for typing the manuscript.

Introduction: The nature and scope of the problem.

A follow-up study of a Head Start population implies an evaluation of pupils, programs or some combination of both. The designer of such a study has an option to examine those variables in macroscopic or microscopic fashion in the sense that he can choose to examine for the presence or absence of effects in grossly dissimilar populations (experimental and control groups), or he can choose to relate particular kinds of effects to particular kinds of programs and pupils. The latter design is capable of giving more powerful answers to the questions raised by the theoreticians and experimentalists, but it can be rather restricted in its capacity to generalize. The precise design and, careful sampling required of the microscopic approach dictate a need for extremely large populations to fill the several cells which would result if meaningful generalization is a goal of the investigation. Thus, only a study of national scope could hope to fill the cells corresponding to a good sample of programs, teachers, and pupils. Lack of a population of this magnitude drastically inhibits the generalizability of any microscopic study.

Macroscopic studies may allow for broad generalization, but the substance of that which is generalized is too often of little significance. Simply searching for the presence or absence of an effect may combine gains and losses into mean changes that obscure the real dynamics underlying the program. A large uni-dimensional program may be examined meaningfully in this manner, but such a program cannot be found in educational fields and is certainly not characteristic of Head Start programs.

The goal of the present investigation is to effect a compromise between those two approaches. We would like to achieve some degree of generalizability; in this case we mean a capacity to generalize to the Head Start program and children of Washington, D.C. To the extent that Washington is representative of urban centers providing Head Start to the children of the ghetto, one would be in a position to make further statements about these kinds of programs. It is not, however, our intention to make this leap, so that we will restrict ourselves to simply the Washington situation. Consequently, our sampling was far ranging and thin, although with certain restriction intrinsic to Washington. These

will be discussed below, but for the present an example can be given. Our population attended summer 1966 Head Start which was jointly funded by OEO and Title I, ESEA. Consequently, each of the 48 Centers established by the Washington, D. C. Board of Education for the summer program was located in an elementary school designated as a Title I target school. This meant that a very large majority of the pupils attending the school were from families meeting the Title I poverty criteria. It also meant that the neighborhood serviced by that elementary school is, from the socio-economic standpoint, very low and homogeneous. Finally, it means that the community and the school is better than 98% Negro in composition. It was our intention to identify the sample in Head Start and follow them in the kindergarten offered in the same school. The eleven sample Centers had a total enrollment of approximately 900 children during the summer. A search of the kindergarten classes offered by these same schools and in which the Head Start sample children should have been enrolled, revealed only 340 Head Start graduates. Clearly there is a great deal of transiency of some kind in this population. Our generalizations are therefore, restricted to a population somewhat more stable than the majority of the Head Start population. The control group was also selected from the kindergarten enrollment, but their rate of transiency can only be guessed. If their rate of transiency is the same as the rate of the Head Start population, then it can be assumed that this control group is made up of about one third stable families and two thirds transients who came from other sections of the city. Thus, our control population may be somewhat less stable than the Head Start population. It is unlikely that the control group has a lower rate of transiency than the Head Start population so the least that can be said about them is that they have either the same degree of family transiency or a greater degree of transiency than the Head Start population. Although this will reduce the internal validity of the present design to some extent, the major impact of these facts is in restricted generalizations. We can draw conclusions only about Head Start children who remain in the neighborhood at least long enough to attend kindergarten in the same building in which they attended Head Start.

Generalizability is not, however, the only goal. With a large enough population, we could seek to relate the variations in output with

several kinds of independent measures. Those measures of greatest interest include the nature of family structure from whence the child comes, the nature of the Head Start teacher's behavior toward the child, and the program means and goals guiding the teacher. On the dependent side, we would be interested in cognitive-verbal measures, skills in dealing with stress producing situations such as test-taking, social-emotional measures, and academic achievement in public schools following the Head Start experience.

However, the examination of several of these microscopic variables in the context of a study large enough to allow for interesting generalizations would require an extensive sample and research team. The compromise arrived at in the present investigation was to restrict the measurement to the least number of theoretically important variables and maintain the largest sample possible. In order to achieve this goal several categories of variables were excluded from this study. First, all measures of the family were excluded, except for a sampling of the socio-economic status. The reason for this is that interview material and survey research methods yield little of value in predicting behavior of preschoolers. Observation of parent-child interactions, and other family structure variables, while potentially of much greater value in such predictions, are extremely cumbersome and therefore quite inappropriate to relatively large scale data collection projects. Failure to measure these extremely important variables in a follow-up study such as this must result in heightened within-group (error) variances and a consequent obscuring of the true differences between groups. This is the price to pay for generalizability.

Since it is known that the families living in the neighborhoods serviced by the sample schools are quite homogeneously low socio-economically, a sample of these families was taken to confirm the nature of our experimental sample. Our sample does not deviate from the known characteristics of the neighborhoods, so that we can accept their status as described by the Board of Education.

The next category of variables excluded was that of program and educational goals. A large city school system will of necessity have a great variety of teachers translating the standard curriculum in a variety of ways. Rather than searching for curriculum similarities and differences, it was decided that a more powerful source of variance in child behavior

is a description of the teacher's actual behavior toward the child. Since, as we will describe below, we are interested in the verbal behavior of the child, a measure of verbal interaction between teacher and child was chosen as an indicator of the input dimension of the Head Start program. Unfortunately, a major complication in the orderly process of this investigation occurred, requiring a shift in the research design.

This project was funded by OEO during the spring of 1966, and preparation for work was begun at that time. However, because of certain complicated fiscal arrangements between the granting agency and the University, funds for the work were not released for expenditure until the last week of Head Start operation at the end of the summer. Personnel to be trained for observation could not be hired until that date, nor could selection of the sample. Consequently, the complete set of measurements on the total sample were taken after the Head Start experience. It was for this reason that the sample was selected from the population of Head Start graduates attending kindergarten, rather than from the population in attendance during the summer program. Since our pretest measures were taken during the fall of 1966 (and since it was not possible to observe pupil teacher interaction in the kindergarten class), this study includes no measures of the educational input to which dependent measures might be referred. Our attempt to be representative of the Head Start classes in our sampling does, however, allow us to speak of the effects of the kind of program a large city developed for its summer classes, on the children of the lower income groups who remained in their neighborhoods to attend kindergarten in the same schools.

On the dependent side, our interests were in estimating the nature and extent of change in the Head Start child's skill in dealing with situations of personal evaluation. We reasoned that a valid reflection of the reduced sense of self worth that is assumed to be a part of the child of poverty, particularly as expressed in the tendency to school failure, is the reaction the child has to being tested. A reduced sense of self worth means, at least logically if not psychologically, an

anticipation of failure and a consequent heightened fear of testing situations. It was our intention to assess the general level of anxiety displayed by the child in the Head Start situation and the specific anxiety displayed in the testing situation. Since we were unable to carry out any observations of the child in either the Head Start class or the kindergarten class, our estimates of these properties of our sample are restricted to data collected in the testing situation and to judgments made by the testers and teachers. Several methods were devised to utilize standard test scores as estimates of test anxieties.

In any such follow-up study, a control group design is a desirable, but not necessary approach. Matching control and experimental groups has such serious problems that Campbell and Stanley for example (1966) indicate a strong preference for randomized procedures and abandonment of matching whenever possible. In the present instance, randomization was not possible since we could have no control over the admissions procedures of the Head Start staff. Further, since it is not possible to measure any of the non-child independent variables (family structure, educational input or pre-kindergarten experience of the control group), matching could occur on a restricted set of variables. We could ensure that the control and experimental child came from the same neighborhood and socio-economic level as the experimental child by matching for kindergarten class. This would at the same time ensure roughly comparable kindergarten experiences. Age and sex could also be matched. However, it is not possible to match or measure the children on dependent measures prior to the Head Start experience. The logic of comparison between experimental and control is somewhat impaired by this situation, although patterns of similarities and differences between these groups might very well compensate for unmeasured differences between them prior to the experimental treatment. In fact, a comparison of the kind of changes which take place over time within each of the two groups, can be reasonably made without reference to the beginning level of performance for either of them. The question to be asked of the data under these conditions is: what is the impact of kindergarten experience for children with and without

Head Start experience? Clearly, it is helpful, but not necessary to have experimental and control children matched on dependent measures at the beginning of Head Start, if this question is asked.

The sample.

The total enrollment roster of 7500 children for the summer 1966 Head Start program was obtained from the Board of Education. These children were distributed across 48 Centers enrolling from fewer than 50 to more than 200 children in each. The Centers were located in public school buildings in the target areas designated by the Title I, ESEA criteria. Children from the immediate neighborhood attended each school.

A sample of eleven Centers were selected from the total to reflect the distribution of the number of classes in each Center. Thus the sample Centers had the same distribution of size of classes as the total sample. The 1775 children enrolled in these sample Centers constituted the pool of experimental subjects for this study. This sample of children also reflected the age distribution of the children in the total population. If we divide the total population into an older group (those who would be five years old by Nov. 1, 1966 and therefore eligible to enter kindergarten after Head Start), and a younger group (those too young to be eligible to enter kindergarten in the fall following their Head Start experience), the older group constitute 54% of the total enrollment and the younger 46%. In the sample Centers, the older group constituted 53% and the younger 47% of enrollees.

Older group

In order to select the sample of experimental children, it was necessary to identify those who actually entered kindergarten. It was expected that many children would move out of the city before school began, that many would not attend kindergarten in the same neighborhood school in which their Head Start class was held, and that many would not attend kindergarten for a variety of reasons. This was, in fact, the case because after careful examination of the kindergarten classes into which the children of the eleven experimental Centers should have been enrolled, only 340 Head Start children were identified.

This group then constituted the pool of subjects from which the sample was selected. It is not possible to determine where the other Head Start children are without examining the enrollment lists in each of the 140 elementary schools in the District of Columbia. Our experimental sample is limited simply to those children who are now enrolled in the kindergartens serving the neighborhoods in which they live and in which their Head Start class was located. It is not possible to estimate the similarity or dissimilarity between those Head Start children who did get into our sample and those who did not because the latter could not be found. Our conclusions in this investigation must be tempered by this fact.

Of the 340 Head Start kindergarten children, a total of 112 were selected to be experimental group. These were selected so that we obtained an equal number of boys and girls, and so that the distribution size of Centers in the total population was reflected in the sample.

A control group was established by selecting a match for each child in the experimental group. This means that for each Head Start child, a non-Head Start child enrolled in the same kindergarten class and of the same age and sex was selected as his control. Thus, a total of 112 control children was selected.

Younger group

Of the 820 children in this group, 160 were selected to reflect the proportionate contribution of the size of their Centers to the total distribution of size of Centers.

The parents of each of these children were telephoned to establish the status and eligibility of the children. Those who were enrolled in a pre-school, day care center, or other educational institution, were excluded from the sample. Those whose parents could not be found were also excluded. This left 90 children who were eligible to be included in the sample of younger children. In each case, parental permission to have the child tested by the project staff was secured.

In order to establish the comparability of the experimental and control groups on socio-economic measures, an interview yielding the following data was administered to both groups: 94% of the mothers and 71% of the fathers of the experimental children were living at

home with the children whereas 100% of the mothers and 80% of the fathers of the control children were living at home. The mean number of children in the experimental families is 3.1 and in the control families is 2.4. The mean highest grade completed is 9th grade for all parents in both groups. The mean total annual family income of the experimental group is $4300 and for the control group, this figure is $5100. Although both groups are at poverty levels, it is apparent that neither represents the hard core of the poverty population and that the control group is slightly higher on the economic scale than the experimental group. However, income level on either the Binet or the Peabody test did not correlate greater than zero, so that such differences in the experimental and control population can be considered insignificant for the present purposes. We feel justified in assuming comparability between these groups.

The measuring instruments.

1. Stanford-Binet, revised, 1960 edition. This instrument was used because it provides a broad, reliable measure of cognitive functioning. It is a relatively high predictor of academic success and it is used as a base against which we wish to evaluate the instruments described below. Finally, the face sheet of this measure contains a thirteen item rating scale of the test-taking behavior of the testee which has been found to be predictive of total IQ and performance in pre school for similar populations (Hess, 1966). Test taking anxiety, a variable of considerable interest to the present study, is measured by several items on this face sheet and by at least one factor which emerges from a factor analysis of the items.

2. Peabody Picture Vocabulary Test. This instrument has been used in several studies of Head Start children with similar results: PPVT test scores are typically lower than Binet IQs and do not correlate as highly with Binet scores when compared to data collected from normative(white, middle income) children (Berger, 1966; Burnett, 1964-65; Eisenberg, 1966; Harding, 1966; Ozer, 1966; Rice, 1967). This might be because the Binet has less of a verbal content at the younger age levels. It might also be, however, that (as argued by the present writer (Cline et al, 1966) the PPVT tends to inhibit the performance of children who have a distaste for being measured. A receptive language task such as the Peabody, which allows only a single right or wrong answer to each item, may prove too restrictive to children who feel uncomfortable in a testing

situation. This would be particularly true for those children who are, in fact, verbally facile as well as anxious in the testing situation. Such an expectation is consistent with the literature on test taking anxiety (e.g., Sarason, et al, 1964, Zigler and Butterfield, 1966). We have predicted, and previously found (Cline, et al, 1966), greater discrepancies between the Binet and the Peabody for children above their group Binet median scores, indicating at least that the discrepancy between these measures is not limited to reduced verbal content of the Binet at these age levels. It should be expected that the Head Start experience would be reflected in the nature of the discrepancies between Binet and Peabody scores, if the Head Start experience does contribute to a child's skill in handling himself in a testing situation.

If the receptive form as well as the verbal content of the Peabody is an important source of variance in this kind of population, then holding the content constant and transforming the test to an expressive form, should have a significant effect on the performance of the child with a distaste for being tested. Such an expressive form has been developed by the present writer, and put through several revisions. Its current form utilizes plates 25 through 62 of the Peabody test, with procedures for administration and a scoring system that have been used several hundred times.(see appendix) It has not however, been used on a kindergarten population. To this point in the development of the instrument, children generally score higher on the expressive (hereinafter referred to as the E form of the Peabody) than the receptive (hereinafter referred to as the R form of the Peabody), and children with higher Binet scores acquire greater discrepancies of the E over the R. However, it is not clear that such predictions would be made for an older kindergarten group. We simply predict that the E form is less inhibiting than the R and that this relationship should discriminate between groups receiving and not receiving a Head Start experience if such an experience is effective in increasing the confidence a child has in taking tests.

3. Behavior Inventory. This instrument is a 50 item rating scale developed by Head Start to assess the social-emotional status of pre-school children. It is used by the teacher to rate the child. Preliminary work with this instrument indicated that a factor structure emerged from its analysis which contained at least one factor which appeared to be conceptually related to the notion of test taking anxieties. Consequently, the measure was included in the present investigation in an attempt to discover such a usable factor in the analysis of our data.

4. Metropolitan Reading Readiness Test. This is the measure administered by the school system at the end of the kindergarten year to assess the readiness of the child for first grade. We shall use these scores as the measure of academic success.

The evaluational hypotheses follow from the expectation that the Head Start experience yields improved performance on the Binet and the several forms of the Peabody (although this is based on the unproved assumption that control and experimental children were at comparable levels on these instruments prior to Head Start); that the Head Start experience is associated with reduced discrepancy scores, greater gain scores over the kindergarten period on the Peabody R and E forms, better academic achievement, and better social-emotional adjustment as revealed by the Behavior Inventory.

The evaluation design.

The original plan called for the start of the testing program during the first week of Head Start. Since the project was not activated until September, this was impossible. The sample was selected by late September, so that testing could not begin until then. The revised design called for pre testing as early in the school year as possible, post testing after a minimum f four months, and the administration of the Behavior Inventory at the same time as the post testing. The reading readiness test is administered by the school system to all kindergarteners early in May so that this date became the end point of the testing program.

The pretest battery included the full scale Stanford-Binet, Peabody-R and Peabody-E. Binet testers were also instructed to fill out the face sheet for each child. Binet testers were hired from the ranks of those who had completed a graduate level course in intelligence testing, who had experience under supervision in administering the Stanford-Binet, and who administered the Binet to at least three low income Negro preschoolers under the supervision of the staff director of testing. All Binet testers were white, nine were women, two were men. All followed the same set of directions and procedures.

The 18 Peabody testers were undergraduate Negro women (except for one white woman and one Negro male) none of whom had administered any sort of test to children before. All were trained by the project staff director, Half the testers were trained to administer the E and half the R form. In order to assure the reduction of tester bias, the following procedures were adopted: All children were randomly assigned to testers; no tester could test the same child twice with the same instrument or with different instruments;

no tester could reveal the score of any child to any other tester (scores were reported only in code) and no tester knew which children were Head Start and which were controls. Some Binet testers did administer the Peabody to some children, but never to the same child to whom he had administered the Binet. Peabody-R tests were scored by the tester, but the Peabody-E tests were scored by two independent judges who adjudicated any differences in the scoring of a child's verbal response in conference. No differences were allowed to remain in the scoring of the E.

In order to eliminate order effects in the administration of these instruments, the following procedure was adopted: The Peabody tests were always administered before or after the Binet. Half the subjects received the E first and half received the R first. This procedure made tester assignments extremely awkward so it was decided to analyze for order effects after the pretest to see if such a balancing was necessary. As reported below, there were no differences between any of the tests associated with their position in the order of administration. Consequently, the balanced order of the Peabody was abandoned for the post testing, although the two Peabody tests were always given before or after the Binet.

Immediately after the pretesting of the older group in school was finished, the younger children (i.e., those who had attended Head Start in the summer but who were not old enough to attend kindergarten) were contacted in their homes. Arrangements were made to test in the home and this was accomplished in the late fall. Only one administration of the battery composed of the Binet and the two forms of the Peabody was carried out for this population. These children were to be used for a follow-up study once they returned to Head Start the next summer.

Table I summarizes the number of tests given to the several populations at the several times during this study.

Table II summarizes the number of boys and girls in the several age and experimental groups.

Results.

Attrition. Although oversampling was carried out in order to compensate for the expected 10% loss of subjects from pre to post, a considerably larger rate of attrition occurred. Each of the dependent measures were given on a

different day, which produced a different rate of attrition for each. These rates varied between 20-30%. In order to determine if this loss represented a selective withdrawal of low or high performing subjects which might serve to bias the post test sample, a comparison of the pre test scores for those subjects who received a post test and those subjects who did not receive a post test was carried out for each of the three dependent measures. Table III summarizes these results, which clearly indicate that the sample was not biased by the loss of subjects. In no case is there any differences between the pre test scores of the subjects who did and who did not take post tests.

Tester effects. The total number of tests administered by each tester was tallied and a simple one way analysis of variance run in order to estimate tester variance. No significant effects were found even though two of the testers produced a mean IQ on the Binet of 77, and two testers produced means of 84 and 86 respectively. Peabody testers were slightly more uniform with means ranging from 82-94. Clearly there is some tester variance (see appendix, Tables II-IV) but the within variance is much too great in this population to give this factor much credence. It would not be wise, however, to generalize these results to other populations or testers.

Order Effects. Mean scores for pre test PPVT-R and PPVT-E in each ordinal position were calculated. No differences were noted indicating that it makes no difference whether the E or R form of this instrument is administered first. The Binet was administered to some children before and some after the Peabody tests but never between the two. No differences in Binet scores related to ordinal effect was noted. All instruments within the battery were administered to each child on separate days but within a week of each other.

Binet results. Table IV summarizes the pre and post test Binet IQs for the several populations. The median IQ for the total population is 89.5 and in subsequent analysis, all S's are sorted in High Binet (90 and higher) and Low Binet (89 and lower) groups. This split is appropriate because we wish to know if the discrepancy between the Peabody scores and Binet scores is the same for high and low Binet scorers.

In order to analyze these data further, an analysis of variance with unequal N's was carried out on the pre test Binet scores against Sex (A), Experimental Status, Head Start/ Control (B), and High (above median)/ Low (below median) Binet IQ (C). This analysis is summarized in Table V. Clearly the only significant effect is (C) which is simply a confirmation of the high/low Binet split. Most important is the non-significance of (B), Experimental status. At the beginning of kindergarten there were no differences between Head Start and control group children on Binet IQ, and no differences between boys and girls. If any differences existed as the result of Head Start, they have disappeared by the first month of kindergarten. On the other hand, the experimental and control children were not measured prior to Head Start and it is conceivable that the Head Start children were below the controls at that time so that the result of the summer program was to raise the Head Start children to the level of the controls. This point cannot be evaluated, although it seems unlikely. However, if the summer did produce meaningful gains for the Head Start children, this should be reflected in some pre test advantage over the control children. If this is not true, it might be expected that the momentum generated over the summer, which brought the Head Start child to the level of the control child (if in fact they were not equal before Head Start) should carry the experimental child forward in kindergarten. The analysis of the rest of our data should speak to this issue.

An analysis of variance of Binet gain scores against Sex, Experimental Status and High/Low Binet pre is summarized in Table VI. The only significant effect is (C) High/Low Binet indicating that the children with high pre test Binet scores show significantly different gain scores than the children with low pre Binet scores. However, further examination of these results suggests that this might be a spurious factor. The mean gain (mean of the distribution of gains) for the high Binet group is minus 2 IQ points, and the mean gain for the low Binet group is plus 5 IQ points. A loss on post test scores for a high group and a gain on post test scores for the low group is precisely what is expected if the true mean gain is zero. This is the regression phenomenon in which the prediction of post from pre includes an asymetrical distribution of post scores at the upper and lower extremes. It would be dangerous indeed to refer to these results as an instance of true change over time.

Of greater importance is the result that both experimental and control groups were identical in their failure to show gains over the kindergarten year

(1.9 and .7 IQ improvement respectively). Sex differences in gain also failed to appear.

PPVT-R Table IV summarizes the pre and post PPVT-R IQs for the several populations. TableVII summarizes the analysis of variance of the pre scores and indicates that Head Start children were not different from control children at the beginning of the kindergarten year on this instrument. An analysis of pre-post (repeated measures) scores indicate that a significant improvement in PPVT-R scores is apparent ($F = 23.08$, df 1/164, $p < .001$). Table VIII summarizes the analysis of variance of these gain scores and indicates that there are no differences in the several groups in amount of gain. That is, boys and girls, Head Start children and control children, high Binet children and low Binet children, all gained about 6-7 points on the PPVT-R over the kindergarten year. In order to examine this finding for regression effects, the total population was further split at the median pre PPVT-R score (85 IQ), and the high and low groups examined for their respective gains. The high scorers show no difference between pre and post (-.3 IQ points) whereas the low scorers show a gain of 11.4 IQ points. It is not clear whether this represents a regression effect or a differential impact of the kindergarten year on the low compared to the high PPVT-R children. The total distribution is displaced so far downward that the lower end contains a disproportionate number of extreme scores, whereas the high group contains very few extreme scores when compared to the normative group. Under these conditions, one would expect greater regression to the mean for the low than the high group, but this is not established one way or the other in these data. Correlates of PPVT-R gain scores are reported below.

PPVT-E Table IV summarizes the pre and post scores for the several populations. Table IX summarizes the analysis of variance of the pre scores against sex, experimental status, and high/low Binet. Significant main effects occur for sex (boys are higher) and high/low Binet (high Binet scores are associated with high E scores), but there are no effects attributable to the Head Start experience. A repeated measures analysis revealed no significant change in the E from pre to post, but an analysis of the gain scores (Table X) indicates that high and low Binet scorers show differential gains on this form of the PPVT. Thus high Binet children show a mean gain of 2 points on the E and low Binet children show a mean loss of 2 points. Of greater significance for the present purposes is that there are no differences associated with the Head Start experience.

In order to examine for regression effects in these data, the E distribution was split at the median and the gain scores for these above and below were separately computed. High children show a mean loss of 5.07 points and

low children show a mean gain of 4.88 points. As with the Binet results, these data look very much like the artifacts of regression rather than true changes in expreme populations.

Table XI summarizes the intercorrelations between the Binet, R, and E measures.

Metropolitan Reading Readiness Test. This instrument was administered to all children in each of the sample classes and total class scores were made available to the present investigation. These data were handled in two ways. First, the mean total MRRT scores for the experimental and control children were tallied and analyzed. This allows a comparison of these groups to each other and to the national norms established for the instrument. However, we also wish to know the position these groups have in their class with respect to reading readiness. That is, independent of absolute level of achievement, are Head Start and control children typically at the top, middle or bottom of their classes. Consequently, the total distribution of each class was established and each of the sample children was assigned a score to reflect his rank in a three position (top third, middle third, bottom third) distribution of scores. Tables XII and XIII summarize the analysis of variance of the MRRT total scores and the MRRT class ranks respectively. In both instances a significant main effect is found for sex (girls score higher than boys and are typically at the higher rank positions in their classes than boys) and Binet IQ (in a system in which rank 1 is the bottom third, rank 2 is the middle third, and rank 3 is the top third, the mean rank for boys is 1.86, for girls 2.26, $F = 15.8$, df 1/181, $p < .01$; mean rank for high Binet IQs is 2.41 and for low Binets IQs 1.72, $F = 45.23$, df 1/181, $p < .001$; mean rank for Head Start children 2.03, for controls, 2.10 (ns)). Clearly, if Head Start generated momentum in these children, it was dispelled by the end of the kindergarten year.

Behavior Inventory. This 50 item rating scale was filled out by 18 teachers describing 800 children including the present sample population at the end of the school year. These data were factor analyzed (principal components, varimax rotation) and a three factor structure emerged which is both interesting and usable (see appendix). Factor I has been labeled Withdrawing, Inhibited, Casper Milquetoast. Factor II has been labeled Hostile, Impulsive, Low controlled, Donald Duck. The third factor has been labeled Non-persistence in Problem Solving.

Each sample child was rescored by summing his scores on the items loading on each factor, and three factor scores were then recorded for each child. The analysis of variance of these scores are given in Tables XIV, XV, XVI.

Factor I significantly differentiates high and low Binet children, but no other main effects or interactions are noted. This means that low Binet children tend to be judged by their teachers as withdrawn and inhibited, and high Binet children are not so judged by their teachers. However, at the end of the school year there does not seem to be any differences between Head Start children and control in their teachers perceptions of them on this dimension.

Factor II reveals no significant main effects but it does show an interaction between sex and high/low Binet. Teachers see low Binet boys and high Binet girls as more hostile and impulsive than high Binet boys and low Binet girls. Most of this effect seems to be the result of the very high impulsive scores given the high Binet girls by their teachers.

Factor III shows a significant sex effect, high/low Binet effect and an interaction between the two. Thus, girls are more persistent than boys in problem solving, high Binet children are very much more persistent than low Binet children, high Binet boys are more persistent than low Binet boys, but there is no difference in the persistence of high and low Binet girl

Binet Face Sheet The thirteen items constituting this scale were factored and four interesting and usable factors emerged. (see appendix) These factors have been named: I Emotional dependence; II Persistence; III Social skills and social confidence; and IV Distractibility. Analysis of the pre test factor scores are summarized in Tables XVII-XX.

Factor I shows just one effect: High Binet children are less dependent than low Binet children. Again the same pattern with respect to the Head Start experience emerges. Binet testers, who did not know the experimental status of the children, were unable to distinguish between Head Start and non-Head Start children in their dependent behavior while taking the pre test, although the testers could distinguish between the high and low IQ children on this dimension.

Factor II the persistence measure, distinguished between high and low IQ children, and none other. Factor III functions in the same manner. Factor IV, the distractibility measure has a significant main effect in the high/low Binet score as well, but there is also a significant interaction between sex and experimental status on this dimension. Experimental

boys are more distractible than control boys; control girls are more distractible than experimental girls.

Discrepancy scores. There are two discrepancy scores of interest: Binet PPVT-R differences and PPVT-R-PPVT-E differences. In both instances we expect the PPVT-R to show lower scores than the other two, although this effect is not expected to be the same for all sub-groups in our populations. We are particularly interested in the nature of these discrepancies at the end of the Head Start experience for Head Start and control children, for boys and girls, and for children above and below the median Binet IQ.

Table XXI summarizes the analysis of variance of pre Binet-R discrepancies. All discrepancies are in favor of the Binet over the Peabody. Girls show a greater discrepancy than boys, and high Binet children show a greater discrepancy than low Binet children. Thus, girls average 7.5 points higher on the Binet than the Peabody, whereas boys average 2.4 points higher on the Binet than the Peabody. High Binet children average 9.5 points higher on the Binet than the Peabody, whereas low Binet children show no difference in their scores. If these discrepancies tell us something about the differences between boys and girls, and between high and low Binet children, they do not tell us about any differences that might exist between children who went to Head Start and those who did not, since there are no significant effects for experimental status.

Table XXII summarizes the analysis of variance of the post Binet-R discrepancies. After a year in kindergarten girls still do better on the Binet than the Peabody to a greater extent than the boys. However, the distinction between the high and low Binet children on this discrepancy measure has disappeared on postest. This results from the fact that the high Binet children (who scored significantly better on the Binet than the Peabody on pre test), lost a few points on the Binet and gained a few on the Peabody over the year. The low Binet children (who showed no difference between their Binet and Peabody scores on pretest) gained a few points on both tests over the year. The difference between their difference scores was thereby reduced. However, this change over time is hard to interpret since so much of the pre-post gains scores for these instruments might be artifactual. Of greater significance to the present work is that Head Start children are indistinguishable from non-Head Start children on pre or post testing. This is particularly important since the pretest distinction between high and low Binet discrepancy scores has constructive validity.

In order to examine the nature of the Binet-R discrepancy measure, these scores were compared to the other data in the present study which bear relevance to test-taking attitudes. Thus, it is expected that if the Binet-R discrepancy is the result of an inhibitory effect generated by the form of the Peabody test, then this discrepancy measure would correlate with Factor I and II of the Behavior Inventory (withdrawn and impulsive, respectively), and with Factor IV of the Binet Face Sheet (distractibility). However, no significant correlations were found among these variables, indicating that the discrepancy score may be reflecting dimensions of the child not recorded by the teachers or the testers, while at the same time it does differentiate between high and low Binet children to the same extent that the teachers and testers make this differentiation.

Tables XXIII and XXIV summarize the analysis of variance of the pre PPVT E and R discrepancies. Here, a different picture emerges than the findings for the Binet-R discrepancy. The only significant effect is across high/low Binet scores, but in this case, the low Binet children do better on the E than the R, whereas the high Binet children do better on the R than the E. There are no effects attributable to the experimental status of the children, and there are no sex differences in these data. On post test, this effect disappears. This is due primarily to the change in discrepancy score of the low Binet children. At the end of the year, these children have improved on the R to the extent that the differences between their scores on the two tests, and the difference between the differences for the high and low Binet children have disappeared. This change in the magnitude of the discrepancy score due to a change in the scores of the low Binet children is the reverse of the effects noted for the Binet-R discrepancy. There the changes were related primarily to variations in scores of the high Binet children. In this case the changes are due to variations in the scores of the low Binet children.

In order to examine the content of the E-R discrepancy, correlates in the factor structures of the Behavior Inventory and the Binet Face Sheet were sought. However, as in the case of the Binet-R discrepancy, no significant correlations were found.

Teacher and school effects The failure to find any effects attributable to experimental status may be the result of a masking effect due to differential performance of Head Start and control children in different schools or in different classes. It may be that teachers or schools differ systematically in their treatment of Head Start children compared to their treatment of the control children. Consequently, analyses across teachers and school were run using reading readiness scores and Behavior Inventory factor scores (the measures used by teachers or most strongly influenced by teacher behavior with respect to the children) as the dependent measures. Both teachers and schools showed significant main effects against reading readiness scores (teachers: $F = 6.96$, df 18/156, $p < .001$. schools:: $F = 8.84$, df, 13/161, $p < .001$), indicating that it mattered greatly to a child's academic achievement level which school he attended and which teacher he had. In order to determine if the Head Start children were differentially affected by the school or teacher, interactions between experimental status and school or teacher were examined. No significant interactions were noted for these variables. Clearly, some teachers are associated with greater reading readiness scores than others (the range of class percentile is 20-82), but there is no evidence that the kindergarten teachers were producing different levels of academic achievement in Head Start graduates than in non-Head Start children. (Tables XXVII - XXIX)

Although it is not part of goal of the present investigation, it is of interest to go one step further in identifying the sources of variability of children in reading readiness at the end of the kindergarten year. We wish to determine if this variability is the contribution of the teacher or the children. Thus, if there is some selective assignment working so that children with greater reading potential are assigned in some manner to some teachers, then the variance across teachers at the end of the year is attributable to the children and not to the teacher. (It is hard to imagine how such an assignment system could work at the beginning of kindergarten, although it would start to work the first time the children were re-assigned to teachers. However, such a difference across teachers at the beginning of the year in these schools could occur by chance.) The only measure taken at the beginning of the kindergarten year which might be relevant to this issue is the Binet. Consequently, a simple one way analysis of variance of teachers against pre Binet scores was run. The $F = 2.48$, df 18/196, $p < .01$. This indicates that there is significant variance across teachers on

the Binet, although closer examination (Newman-Koules test of multiple means) indicates that almost all of this variance is contributed by a single teacher whose experimental and control children at the beginning of the year had a mean IQ of 74.3. The class percentile in reading readiness for this teacher at the end of the year is 42.0 which is well within the normal limits for this population. However, it is clear that the mean pre Binet IQ in 18 of the 19 classrooms are equivalent to each other and that at the end of the year there is very great differences across these classrooms. In order to substantiate this point, an analysis of covariance was run with the pre Binet scores as the covariate. With the pre Binet scores covaried the F equals 2.68 df 18/196 p is less .01, indicating that there are significant differences across teachers on reading readiness. Since the substance of differences between teachers cannot be examined further with the data of the present investigation, we shall report no further results of this analysis. Clearly, teachers, as well as pupil characteristics, contribute meaningfully to reading readiness achievement at the end of the kindergarten year.

Teachers rated the children in their classes on the Behavior Inventory within a week of the administration of the reading readiness tests. In order to determine if there are differences across teachers in the assignment of such ratings, and if there are interactions between teachers and experimental status, an analysis of variance of each factor score was run. Tabels XXVII-XIX summarize these analyses. In each instance there is a significant main effect for teachers. Factors I and III also show a significant effect for sex (boys are judged as more withdrawn than girls, and girls are judged more persistent than boys). Of interest, however, are interactions between experimental status and teachers on factors I and III. Visual examination of these analyses reveals that there is no systematic direction to the interaction, which are weak at the most. In both instances the largest source of the effect comes from a single teacher, who judged Head Start children as more withdrawn and lacking in persistence than control children. This teacher appears to be different from all other teachers in these judgements. It would be difficult to attribute any further significance to these interactions. Although it is not possible to ascribe these post characteristics to teachers, since there are no pre Behavior Inventory ratings with which to make comparisons, it is clear that at the end of the year, children differed on these characteristics depending upon the teacher with whom he has been associated. This effect might, of course, reflect a teacher responsible bias rather than real differences between children, but this analysis is beyond the purposes of the present study.

Interrelations between instruments. We are interested in certain interrelations between the measures for a number of reasons. First, although several of the measures differentiate between children, albeit not Head Start and control children, it is necessary to know if these measures are related to criterion measures of theoretical importance. Thus, it is important to know if the measures we have taken predict reading readiness with efficiency. If so, the interpretation of the results reported above is clarified.

Next, it is clear that the only measure which shows significant improvement over the year is the PPVT-R. In order to interpret this, the correlates of such a gain would be helpful.

Finally, the discrepancy between the Binet and the PPVT-R scores are very variable, and of theoretical significance. This discrepancy score is not related to several variables which serve as construct validity criteria. Consequently, a further investigation of its correlates is necessary to help interpret these data.

Each of these interrelations was carried out by way of a stepwise multiple regression analysis summarized in Tables XXX, XXXI and XXXII. In predicting scores on the MRRT, it is clear that the set of predictor variables (the Binet IQs, Peabody IQs, Binet Face Sheet factor scores, and Behavior Inventory factor scores) account for a very large portion (66.5%) of the variance of the criterion variable. The variables with the largest beta weights (1.02 - 2.53) in this set of predictors are the Peabody scores, the Binet IQ, and Behavior Inventory factor III (Persistence). Our purpose here is not to make predictions of the criterion variable, but to demonstrate that the measures in our battery are meaningful with respect to reading readiness. It is clear that this is true, particularly for the Persistence score and the Binet IQ.

The PPVT-R gain score is of interest because it is the only gain to reach significance. Here, examination of the iteration sequence shows that the two discrepancy measures constitute the best set of predictors accounting for 37% of the variance probably because both contain the PPVT-R as a component. As reported above, low R scores (and consequently high discrepancy scores) are associated with high R gain scores. The full battery including those first two variables account for 41% of the variance of R gain scores indicating that there are few correlates of any meaning for this variable. This is consistent with

the notion that PPVT-R gain scores are largely the expression of a regression phenomenon rather than meaningful gains.

Examination of the iteration sequence for the Binet-R discrepancy shows that the first three variables constitute the best set of predictors, accounting for 47% out of the total of 57% of the variance accounted for by the full battery. Once again, the variables with construct validity for this variable (the Behavior Inventory and Binet Face Sheet) do not contribute meaningfully to its variance. The E-R discrepancy, which is the first variable in the iteration, and the PPVT-E, which is the second, both should reflect the same phenomenon as the Binet-R discrepancy, viz., expressive tasks such as the PPVT-E and Binet, represent lesser demands for the populations used in the present investigation. Operating in consort, these variables seem to relate to each other although external validity is still lacking.

Discussion.

The most important finding around which this discussion must center is the failure to discover any effects associated with the Head Start experience. This includes any effects immediately after the Head Start summer, and those longer term effects of almost a year later. The expectation that a summer of Head Start may serve to inoculate the child against the sources of academic failure and, therefore, show up sometime after the injection to protect when the disease attacks, does not appear to be fulfilled in these data. The first issue raised by these findings is whether, in fact, the battery of measurements is sensitive to the kinds of changes which might have been produced by Head Start. This is an unanswerable question, of course, but an approximation of an answer can be made. Clearly, many measures of social-emotional skills were not included, and many cognitive dimensions were not tapped. Motivational changes were not completely sampled. Further, many of the school-related measures that were taken in this study are not necessarily included as part of the stated goals of Head Start. Nevertheless, the effects of an educational program must ultimately be felt in academic work. It might well be that this ultimate effect is produced by changing the child's sense of self worth, motivation to learn, sense of mastery over the environment, anxiety levels, and a feeling of comfortableness

in the educational setting. But ultimately, these changes should show up in academic achievement, if some external event (such as the behavior of the teacher is post-Head Start experiences) does not intervene to destroy the effects. We feel justified therefore in using reading readiness scores at the end of the kindergarten year as the criterion for the battery of instruments in this kind of evaluation study. The finding that a multiple regression analysis can be generated which correlates .85 with the criterion variable, is very supportive evidence that the instruments used in the present study are quite relevant to educational output. Further, the finding that both the cognitive and personality measures differentiated between groups of subjects who are known to be different (boys and girls; high and low performers on the Binet), suggests that the instruments are sensitive to important differences between subjects. This underscores the failure to find any differences associated with Head Start experiences.

Our battery of instruments included two measures of the childrens' behavior: Behavior Inventory, and the Binet Face Sheet. In both cases a clear syndrome of behaviors is associated with the Binet IQs and with reading readiness. Children who are persistent in solving problems, relatively independent, not overly timid and withdrawing, and socially confident, are clearly the successful children in school and on the Binet test (zero order correlations between all of these variables are significant as well as their analyses of variance and multiple regression analysis). If these variables can differentiate between those who are high and low on a measure such as the Binet, and can predict relative success in school, it could be expected that an educational program such as Head Start could be evaluated by them as well. It is true that the Behavior Inventory was administered by teachers, and this might serve to obscure or distort difference between children on these measures. That is, teachers might be rating children differentially according to their status as Head Start graduates. In fact, there was no interaction between teachers and experimental status, indicating that teachers were telling the difference between children but not along the Head Start-non-Head Start dimension. Further, the Binet testers who rated the children in the Binet Face Sheet with respect to their test taking behavior, did not know which children were experimental and which were control children. No differences between these groups occurred on the Face Sheet measures, although other differences did occur.

The Peabody Picture Vocabulary Test (R) should be examined further. This instrument has shown two characteristics when administered to deviant populations: it generates IQ scores which are typically below those of the Binet and it produces rather large gain scores. These factors, together with its ease of administration probably accounts for its popularity in evaluation studies of programs such as Head Start. There are, however, several reasons why this may be an unjustified popularity.

An instrument that systematically depresses scores (relative to the Binet) may be perfectly usable if one knows the extent of the systematic depression in order to make the appropriate adjustments. However, in this case, it is apparent that the discrepancy between the Binet and the Peabody is very much a function of the Binet score. The discrepancy is significantly higher for high Binet scores than for low Binet scores. This finding replicates the finding of the present writer on a younger population (Cline, et al, 1966) and indicates that the discrepancy is not simply a matter of the scoring system of the Peabody or the reduced verbal content of the Binet at early age levels. In the previously cited work, the present writer suggested that this discrepancy reflects the inability of test-anxious subjects to cope with the kind of receptive task provided by the Peabody. There is a great deal of variability in this discrepancy, but in the present study, its only correlates are the Binet scores and other discrepancy measures involving the Peabody. It does not relate to any other construct validity criteria in the present study. The low correlations between the Peabody (R) and the Binet (.44 in the present study, compared to .75 as reported in the manual for the Peabody) results in part from the restricted age range of the present population compared to the age range in the standardization population of the Peabody, and in a slightly restricted range of Binet IQs. But there is the clear implication that these two instruments are not measuring the same functions despite the claims of the Peabody test constructors. It is not clear what the Peabody is measuring, although it is clear that it underestimates the higher Binet scores more than the lower Binet scorers.

The large gain scores which seem to be characteristic of this instrument were found in the present population as well. One might be tempted to say that

the consequence of a year in kindergarten produced an improvement in Peabody scores if it were not for another finding of this study. The gain scores on the Peabody are restricted almost entirely to those children who scored below the median score of the Peabody on pretest. Those above the median show no gain. An alternative explanation (to the hypothesis that this is a true experimental finding), is that such a pattern should be found with random changes from pre to post on an instrument which produces disproportionately low scores. This is one half of the regression phenomenon in which there is greater regression to the mean found in the more extreme scores than in the less extreme scores. The large gains shown on the Peabody, appear to be due to the random fluctuations of the very large numbers of scores at the very low end of the Peabody distribution. This explanation needs to be examined in much greater detail than our present purposes allow. It is reasonable to assert, however, that until the regression hypothesis is rejected, Peabody data collected from low income preschoolers must be considered suspect.

In order to examine the suggestion cited above, that the Peabody (R) underestimates IQ scores because of its uniquely receptive nature, an expressive form of the Peabody was developed by the present writer. It was expected that this form (E form of the Peabody) would produce higher scores than the receptive form. This did occur, but only for those subjects who are below the median Binet score. Thus, for low Binet children, the E form is easier than the R, and for high Binet children, the Binet is easier than the R. On post test, both of these discrepancies disappear, primarily because of the large (and perhaps spurious) improvement in R scores. (Binet-R and E-R correlations with PPVT-R gain score are .50 and .47 respectively indicating that large discrepancies are caused by low Rs which produce high R gain scores).

Although little construct validity is available for these discrepancy measures, it is clear that they differentiate between Binet scores. They do not differentiate between Head Start and non-Head Start children, but since it has not yet been possible to demonstrate that they are related to test anxieties, it is difficult to interpret this finding.

A good deal of the comparison between Head Start and control children is based on the assumption of their equivalence prior to Head Start. The only data relevant to this point are the pretest measures taken a month or two after Head Start ended. It would be reasonable to assume that if there are any Head Start effects in these data, it would show up at that time. There are, of course, two reasons why significant gains attributable to Head Start would not show up on testing in October. The first is that six weeks to a month of kindergarten would wash out any such effects. The second is that the true gain of the Head Start children was to catch up to the control children who were more advanced than the experimental children at the time of entrance to Head Start. This latter point is less convincing because if the Head Start children did, in fact, start out at a lower level and simply caught up in the summer, then some kind of reflection of that would be found in one of the tests, or in one of the gain scores. A child who has been able to make large gains in a summer program should be able to exhibit some momentum for a few months at least. The only alternative to this is that the gain over the summer was so tenuous as to be easily dispersed by the beginning of school. Such a gain would be hardly worth striving for.

The hypothesis that a gain might have been present, but washed out by the kindergarten teacher during the first few months of school, is a more reasonable one. If this is the case, however, it means that the Head Start child suffered more at the hands of the teacher than the non-Head Start child. It was to test this hypothesis that the several interactions between teachers and experimental status were examined. None were found to be close to significant. It is clear that the kindergarten teacher does exert a great deal of influence over the performance of the child, and if that influence was negative, it might be great enough to obscure differences between experimental and control children during the first few months. This is a clear possibility that cannot be examined in in the present study. Teacher behavior is an extremely important and, as yet, unexamined variable in this investigation. The answer to these questions can only come from careful observation of teacher behavior from the very first day of school, together with measures of the children taken after Head Start but before kindergarten.

Conclusions

We have attempted to sample from the populations of teachers and children of the kindergarten classes of the Negro ghetto of Washington, D.C. Half of the children in the sample attended summer Head Start preceding kindergarten; half did not. Although not strictly equivalent in the sense of the random sampling model, the two groups of children were drawn from the same socioeconomic stratum. The qualifications concerning equivalence, discussed above, restrict the generalizability of our conclusions to a population which is, perhaps, somewhat less transient than the Head Start population for the whole city. However, in the most limited sense, the findings do represent the probable effect of Head Start on the academic readiness and social-emotional development of economically impoverished children of the District of Columbia.

The sampling of programs was representative of those offered lower income children by the D.C. Board of Education. Even without an in depth analysis of either Head Start or kindergarten programs, it is possible to conclude that no discernible effect of Head Start appeared. Whether this failure lay with Head Start, or whether some kind of stereotyped teacher responses in the kindergartens obscured Head Start effects, is still not clear.

Another conclusion to be drawn is that the variability observed across our several variables was much greater than would be expected by chance. To be culturally disadvantaged clearly has no uniform meaning. The children of the lower income families are much more heterogeneous than is often supposed. It is entirely possible, therefore, that the failure of teachers described above, depends on

their lack of awareness of the range and variability of the skills present in their classes, and the nature of the deficiencies likely to be present in individual children.

A final conclusion, and one that must serve as a warning to researchers in this field, is that scores on reliable, much used instruments such as the Peabody have both more variability and less validity among children of the urban poor than upon the children on whom they were standardized.

BIBLIOGRAPHY

erger, S. I. Development of appropriate evaluation techniques for screening children in a Head Start program -- a pilot project. (Final Report, OEC Contract No. 515, 1966).

urnett, A. Comparison of the PPVT, Wechsler-Bellevue and Stanford-Binet on educable retardates. American Journal of Mental Deficiency, 1964-65, 69.

ampbell, D. T. and Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

line, M. G., Marshall, Judith and Stansbury, Eunice. Stanford-Binet, PPVT and low-income preschoolers: new pitfalls for old tests. (Read at the Annual Meeting of the Eastern Psychological Association, New York City, April 1966).

isenberg, Leon. A. Change in verbal intelligence as measured by the Peabody Picture Vocablulary Test; B. Teacher behavior and attitudes in relation to changes in the children; C. The effects of special perceptual training on the Peabody Picture Vocabulary Test, the Columbia Mental Maturity Scale and a new test, the John Hopkins Perceptual Task; E. Clinical testing of Head Start children. (Final Report, OEO Contract No. 510, 1966.)

arding, John. A comparative study of various Project Head Start programs. (Progress Report, OEO Contract No. 545, 1966).

ice, J. et al. Validity of the PPVT for low IQ children. American Journal of Mental Deficiency, 1967, 71.

zer, Mark. The effects of neurological and environmental factors on the language development of Head Start children: an evaluation of the Head Start program. (Final Report, OEO Contract No. 528, 1966).

arason, S. B., Hill, K. T. and Zimbardo, P. G. A longitudinal study of the relation of test anxiety to performance on intelligence and achievement tests. Monograph of the Society for Research in Child Development, 1964, 29, Whole No. 98 (7).

igler, Edward and Butterfield, E. C. Motivational factors and IQ changes in culturally deprived children attending nursery school. (Final Report, USPHS Grant MH-06809, and OEO Contract No. 2405, 1966.

*Hess, Robert D., Principal Investigator. Techniques for assessing cognitive and social abilities of children and parents in Project Head Start. Final Report, OEO Contract 519. University of Chicago, 1966.

LIST OF TABLES

LIST OF TABLES (cont.)

TABLE I

Number of children receiving each of the measures.

	Pre-Tests			Post-Tests				
	Binet	PPVT-R	PPVT-E	Binet	PPVT-R	PPVT-E	BHI	MMRT
Experiment	112	111	113	87	103	87	104	102
Control	103	100	105	78	83	77	91	92
Total	215	211	218	165	136	164	195	194
Invalid	0	0	0	1	2	0	0	0
Total (usable tests)	215	211	218	164	184	164	195	194

TABLE II AGE AND SEX COMPOSITION OF SCHOOL SAMPLE AS OF OCT. 31, 1966

	EXPERIMENTAL			CONTROL			
AGE (MOS)	BOYS	GIRLS	N	BOYS	GIRLS	N	TOTAL N
79-81	0	1	1	0	1	1	2
76-78	0	0	0	0	0	0	0
73-75	0	1	1	0	0	0	1
70-72	8	9	17	8	9	17	34
67-69	16	13	29	7	13	20	49
64-66	14	12	26	14	12	26	52
61-63	20	17	37	19	15	34	71
58-60	5	4	9	6	4	10	19
N =	63	57	120	54	54	108	228
$\bar{X}$=	65.60	66.24		65.06	66.22		

$\bar{X}$ Experimental = 65.90

$\bar{X}$ Control = 65.64

$\bar{X}$ - Total = 65.78

TABLE III

PRETEST MEANS FOR SUBJECTS
WITH AND WITHOUT POSTESTS

	TEST					
	N	BINET	N	PPVT-R	N	PPVT-E
With Postest	159	89.98	176	84.72	166	27.04
Without Postest	56	89.43	35	87.00	56	28.27

TABLE IV. MEANS AND STANDARD DEVIATION OF PRE AND POST
PPVT-R, PPVT-E and STANFORD BINET

	PPVT - R (IQ)				PPVT - E (Raw Sc.)				Stanford-Binet (IQ)			
GROUP	N	PRE	N	POST	N	PRE	N	POST	N	PRE	N	POST
EXPERIMENT	111	83.53 15.80	103	90.94 15.86	112	27.48 5.84	86	31.20 6.70	113	89.23 14.35	87	90.38 12.43
BOYS	54	85.70 15.26	53	92.96 12.44	58	28.41 5.74	45	32.87 6.36	60	87.82 13.17	44	89.86 9.39
GIRLS	57	83.42 16.34	50	88.80 18.72	54	26.48 5.83	41	29.37 6.67	53	90.83 15.56	43	90.91 15.02
CONTROL	101	85.94 15.76	82	91.27 15.69	105	27.23 6.85	78	29.69 5.96	102	90.52 12.22	77	92.68 14.43
BOYS	55	87.40 16.14	40	90.83 18.18	56	28.16 7.64	42	30.50 5.94	53	89.30 12.20	41	91.51 15.20
GIRLS	46	84.20 15.28	42	91.69 13.10	49	26.16 5.71	36	28.75 5.92	49	91.84 12.22	36	94.00 13.58
TOTAL	212	85.20 15.76	185	91.09 15.74	217	27.36 6.34	164	30.48 6.39	215	89.84 13.37	164	91.46 13.41

The top number in each cell is the Mean; the lower is the Standard Deviation.

TABLE V ANALYSIS OF VARIANCE, PRE BINET

SOURCE	DF	MS	F
Sex (A)	1	76.57	1.07
Experimental Status (B)	1	182.16	2.54
High/Low Binet (C)	1	22426.06	313.316***
A X B	1	217.14	3.03
A X C	1	29.96	0.419
B X C	1	124.73	1.74
A X B X C	1	87.87	1.23
Error	207	71.57	1.00

*** $P < .001$

TABLE VI ANALYSIS OF VARIANCE, BINET GAIN

SOURCE	DF	MS	F
Sex (A)	1	12.47	0.134
Experimental Status (B)	1	55.84	0.601
High/Low Binet (C)	1	1948.07	20.957***
A x B	1	23.77	0.256
A x C	1	0.895	0.010
B x C	1	18.39	0.198
A x B x C	1	139.17	1.497
Error	153	92.96	1.000

*** $P < .001$

TABLE VII. ANALYSIS OF VARIANCE, PRE PPVT-R

URCE	DF	MS	F
x (A)	1	670.369	3.089
perimental Status (B)	1	68.578	0.316
gh/Low inet (C)	1	7008.750	32.293 ***
X B	1	51.098	0.235
X C	1	3.980	0.018
X C	1	74.930	0.345
X B X C	1	13.064	0.060
ror	194	217.037	1.000

* P < .001

TABLE VIII. ANALYSIS OF VARIANCE, PPVT-R GAIN

SOURCE	DF	MS	F
Sex	1	66.84	0.270
Experimental Status (B)	1	6.82	0.280
High/Low Binet (C)	1	1.48	0.006
A x B	1	18.96	0.077
A x C	1	127.90	0.516
B x C	1	67.96	0.274
A x B x C	1	875.32	3.534
Error	163	247.67	1.000

TABLE IX . ANALYSIS OF VARIANCE PPVT-E (PRE)

SOURCE	DF	MS	F
Sex (A)	1	5426453.875	5.453 *
Experimental Status (B)	1	116893.243	0.117
High/Low Binet (C)	1	7412862.643	7.449 **
A X B	1	75143.802	0.076
A X C	1	728691.796	0.732
B X C	1	9974.323	0.010
A X B X C	1	737811.485	0.741
Error	198	995082.508	1.000

* P < .05

** P < .01

TABLE X. ANALYSIS OF VARIANCE, PPVT-E GAIN

SOURCE	DF	MS	F
Sex (A)	1	594721.99	0.478
Experimental Status (B)	1	191322.21	0.154
High/Low Binet (C)	1	6649088.81	5.342*
A x B	1	9832.51	0.008
A x C	1	2707371.86	2.175
B x C	1	357633.15	0.287
A x B x C	1	730438.01	0.587
Error	149	1244687.94	1.000

* P < .05

TABLE XI. Intercorrelations between

Stanford-Binet, PPVT-R and PPVT-E (Pre-Tests)

	PPVT-R	PPVT-E
anford-Binet	.446	.315
VT-R	---	.399

= 212)

TABLE XII. ANALYSIS OF VARIANCE,

Metropolitan Reading Readiness Test : Total Score

OURCE	DF	MS	F
ex (A)	1	1831.247	6.750 **
xperimental Status (B)	1	228.870	0.844
igh/Low Binet (C)	1	14136.379	52.105 ***
X B	1	48.991	0.181
X C	1	272.296	1.004
X C	1	40.570	0.150
X B X C	1	158.369	0.586
rror	167	271.303	1.000

** P <.01

*** P <.001

TABLE XIII. ANALYSIS OF VARIANCE,

Metropolitan Reading Readiness Test : Class Rank

URCE	DF	MS	F
x (A)	1	7.460	15.888 ***
perimental Status (B)	1	0.272	0.578
gh/Low Binet (C)	1	21.333	45.438 ***
X B	1	0.469	0.999
X C	1	0.234	0.499
X C	1	0.394	0.840
X B X C	1	0.068	0.145
ror	181	0.470	1.000

** P< .001

TABLE XIV. ANALYSIS OF VARIANCE, BEHAVIOR INVENTORY

FACTOR I

SOURCE	DF	MS	F
Sex (A)	1	27.62	0.300
Experimental Status (B)	1	133.00	1.445
High/Low Binet (C)	1	2901.68	31.534***
A x B	1	306.21	3.328
A x C	1	65.57	0.713
B x C	1	6.05	0.066
A x B x C	1	27.60	0.300
Error	197	92.02	1.000

*** $P < .001$

TABLE XV. ANALYSIS OF VARIANCE, BEHAVIOR INVENTORY

FACTOR II

SOURCE	DF	MS	F
Sex (A)	1	149.99	1.351
Experimental Status (B)	1	13.93	0.126
High/Low Binet (C)	1	57.02	0.514
A x B	1	129.86	1.17
A x C	1	534.02	4.810*
B C	1	270.30	2.435
A x B x C	1	12.81	0.115
Error	197	111.01	1.000

* P < .05

TABLE XVI. ANALYSIS OF VARIANCE, BEHAVIOR INVENTORY

FACTOR III

SOURCE	DF	MS	F
Sex (A)	1	584.35	6.83 **
Experimental Status (B)	1	99.07	1.158
High/Low Binet (C)	1	1830.89	21.40 ***
A x B	1	17.04	0.199
A x C	1	498.37	5.825 *
B x C	1	2.17	0.025
A x B x C	1	8.10	0.095
Error	207	85.563	1.000

* $P < .05$

** $P < .01$

*** $P < .001$

TABLE XVII ANALYSIS OF VARIANCE, Pre Binet Face Sheet: Factor I

SOURCE	DF	MS	F
Sex (A)	1	54.155	0.506
Experimental Status (B)	1	0.192	0.002
High/Low Binet (C)	1	1129.715	10.563**
A X B	1	3.611	0.034
A X C	1	23.063	0.216
B X C	1	1.150	0.011
A X B X C	1	152.263	1.424
Error	190	106.948	1.000

** P < .01

TABLE XVIII ANALYSIS OF VARIANCE, Pre Binet Face Sheet: Factor II

OURCE	DF	MS	F
ex (A)	1	215.082	2.423
xperimental Status (B)	1	34.950	0.394
igh/Low Binet (C)	1	2695.524	30.373***
X B	1	3.077	0.035
X C	1	170.776	1.924
X C	1	179.081	2.018
X B X C	1	52.128	0.587
rror	189	88.749	1.000

*** P < .001

TABLE XIX. ANALYSIS OF VARIANCE, Pre Binet Face Sheet: Factor III

URCE	DF	MS	F
: (A)	1	296.073	1.984
xperimental Status (B)	1	34.815	0.233
lgh/Low Binet (C)	1	851.218	5.705 *
X B	1	24.297	0.163
X C	1	0.011	0.000
X C	1	247.864	1.661
X B X C	1	16.960	0.114
rror	190	149.210	1.000

P < .05

TABLE XX ANALYSIS OF VARIANCE, Pre Binet Face Sheet: Factor IV

SOURCE	DF	MS	F
Sex (A)	1	288.594	2.867
Experimental Status (B)	1	1.988	0.020
High/Low Binet (C)	1	2794.507	27.765***
A X B	1	681.283	6.769**
A X C	1	1.437	0.014
B X C	1	15.845	0.157
A X B X C	1	14.443	0.143
Error	190	100.649	1.000

** $P < .01$

*** $P < .001$

TABLE XXI . ANALYSIS OF VARIANCE,

Pre Discrepancy Score : Stanford-Binet - PPVT-R

SOURCE	DF	MS	F
Sex (A)	1	1270.576	5.791 *
Experimental Status (B)	1	51.205	0.233
High/Low Binet (C)	1	3969.262	18.092 ***
A X B	1	45.567	0.208
A X C	1	70.263	0.320
B X C	1	4.701	0.021
A X B X C	1	100.456	0.458
Error	194	219.396	1.000

* P < .05

***P < .001

TABLE XXII. ANALYSIS OF VARIANCE,

Post Discrepancy Score : Stanford-Binet - PPVT-R

URCE	DF	MS	F
ex (A)	1	971.742	3.942 *
xperimental Status (B)	1	1.351	0.005
gh/Low Binet (C)	1	144.951	0.588
X B	1	87.854	0.356
X C	1	3.282	0.013
X C	1	15.788	0.064
X B X C	1	1.498	0.006
rror	146	246.534	1.000

P < .05

TABLE XXIII. ANALYSIS OF VARIANCE,

Pre Discrepancy Score: PPVT-E - PPVT-R

URCE	DF	MS	F
x (A)	1	260203.421	0.214
perimental Status (B)	1	282835.970	0.233
gh/Low Binet (C)	1	7616119.177	6.271 *
X B	1	23348.428	0.19
X C	1	367568.504	0.303
X C	1	142279.098	0.117
X B X C	1	1384582.394	1.140
ror	190	1214567.128	1.000

P < .05

TABLE XXIV. ANALYSIS OF VARIANCE,

Post Discrepancy Score: PPVT-E - PPVT-R

SOURCE	DF	MS	F
Sex (A)	1	1307506.654	1.158
Experimental Status (B)	1	814727.086	0.722
High/Low Binet (C)	1	1591418.126	1.410
A X B	1	57485.337	0.051
A X C	1	67567.069	0.060
B X C	1	234.729	0.000
A X B X C	1	4053738.040	3.591
Error	140	1129008.245	1.000

TABLE XXV . ANALYSIS OF VARIANCE
Metropolitan Reading Readiness Test

Sex, Experimental Status and Teacher Effects

SOURCE	DF	MS	F
Sex (A)	1	1791.615	9.534 **
Experimental Status (B)	1	18.318	0.097
Teachers (C)	8	1541.191	8.202 ***
A X B	1	88.178.	0.469
A X C	8	302.824	1.611
B X C	8	86.300	0.459
A X B X C	8	147.922	0.787
Error	80	187.915	1.000

** P < .01

*** P < .001

TABLE XXVI . ANALYSIS OF VARIANCE
Metropolitan Reading Readiness Test:

Sex, Experimental Status and School Effects

RCE	DF	MS	F
(A)	1	2486.229	11.159 ***
perimental Status (B)	1	71.866	0.323
hools (C)	7	2389.778	10.726 ***
X B	1	9.944	0.045
X C	7	242.082	1.087
X C	7	106.022	0.476
X B X C	7	187.709	0.842
ror	115	222.807	1.000

* P < .001

TABLE XXVII. ANALYSIS OF VARIANCE, Behavior Inventory I: Sex, Experimental Status and Teacher Effects

URCE	LF	MS	F
x (A)	1	372.972	5.669 *
xperimental Status (B)	1	62.752	0.954
eachers (C)	8	779.926	11.855 ***
X B	1	1.659	0.025
X C	8	108.869	1.655
X C	8	148.901	2.263 *
X B X C	8	14.433	0.219
rror	101	65.787	1.000

P < .05

** P < .001

TABLE XXVIII. ANALYSIS OF VARIANCE, Behavior Inventory II:

Sex, Experimental Status and Teacher Effects

JRCE	DF	MS	F
x (A)	1	44.307	0.885
perimental Status (B)	1	2.328	0.047
achers (C)	8	1826.836	36.493 ***
X B	1	0.086	0.002
X C	8	33.269	0.665
X C	8	26.190	0.523
X B X C	8	5.400	0.108
ror	101	50.060	1.000

* P < .001

TABLE XXIX . ANALYSIS OF VARIANCE, Behavior Inventory III:
Sex, Experimental Status and Teacher Effects

URCE	DF	MS	F
x (A)	1	647.751	11.907 ***
xperimental Status (B)	1	31.636	0.582
eachers (C)	8	715.034	13.144 ***
X B	1	9.394	0.173
X C	8	116.043	2.133
X C	8	129.956	2.389
X B X C	8	22.833	0.420
rror	109	54.399	1.000

** P < .001

TABLE XXX . Multiple Regression

Criterion Variable - MRRT (Total Score)

Variable	Multiple R	Increase in R^2
Behavior Inventory III	.649	.421
Post Binet IQ	.718	.094
Pre Binet Face Sheet II	.731	.019
Pre Binet IQ	.742	.015
Pre Binet Face Sheet IV	.755	.019
Pre Binet Face Sheet I	.763	.013
Pre PPVT-R IQ	.772	.013
Behavior Inventory II	.776	.006
Post Binet Face Sheet IV	.780	.005
Pre E-R Discrepancy	.782	.003
Pre PPVT-E (T Score)	.803	.033
Post Binet Face Sheet II	.805	.003
Post E-Binet Discrepancy	.807	.002
Post PPVT-E (T Score)	.812	.009
Post E-R Discrepancy	.813	.001
Post PPVT-R IQ	.815	.002
Pre Binet Face Sheet III	.815	.0002
Behavior Inventory I	.815	.0001
	Total R^2=	.665

TABLE XXXI . Multiple Regression

Criterion Variable - PPVT - R Gain Scores

riable	Multiple R	Increase in R^2
e E-R Discrepancy	.568	.323
e Binet-R Discrepancy	.605	.043
e Binet IQ	.612	.007
e Binet Face Sheet III	.622	.012
ehavior Inventory III	.629	.009
RRT Class Rank	.634	.005
e Binet Face Sheet I	.636	.003
e Binet Face Sheet II	.640	.004
ehavior Inventory II	.641	.001
ehavior Inventory I	.642	.0006
e Binet Face Sheet I	.642	.0004
	Total R^2=	.4128

TABLE XXXII, Multiple regression

Criterion Variable - Pre Binet-R Discrepancy

riable	Multiple R	Increase in R^2
e E-R Discrepancy	.408	.166
e PPVT-E	.604	.199
RT Class Rank	.688	.108
navior Inventory II	.714	.036
e Binet Face Sheet III	.728	.020
e Binet Face Sheet I	.7332	.007
e Binet Face Sheet IV	.740	.010
navior Inventory I	.745	.008
havior Inventory III	.753	.011
	Total R^2 =	.568

APPENDIX

A		Instructions for Administration of PPVT-Expressive
B,C,D,E,		Items and Scoring Criteria for PPVT-Expressive
F	Table I	PPVT-Expressive, Reliability Coefficients
G	Table II	Means and Standard Deviations for Stanford-Binet by Testers
H	Table III	Means and Standard Deviations for PPVT-R by Testers
I	Table IV	Means and Standard Deviations for PPVT-E by Testers
J,K	Table V	Behavior Inventory: Factor Structure
L	Table VI	Stanford-Binet Face Sheet: Factor Structure
M		Younger Sample
N	Table VII	Younger Sample: Means, Standard Deviations and Mean Discrepancy Scores for Stanford-Binet, PPVT-R and PPVT-E

PEABODY PICTURE VOCABULARY

EXPRESSIVE FORM

1. Record the child's name, code number, your name, date and school on the answer form.

2. Follow the general administration instructions outlined in the manual (page 6), items 1-6. All children, regardless of age, are given all the items indicated on the answer sheet.

3. Begin the test by saying: I want to play a picture game with you. Turn to Example A, and say: See all the pictures on this page, I will point to a picture and I want you to tell me what it is. Let's try one. Point to the picture of the bed. When the subject makes the desired response, say: That's a good answer and turn to Example B. On Example B, point to butterfly and ask: Now what is this? Do the same for Example C.

4. Turn to Plate 25, but before showing it to the child, say: Now I am going to show you some other pictures. Each time I will point to a picture and I want you to tell me what it is. When we get further along, there may be some pictures you are not sure of, but I want you to try to tell me about it anyway.

5. Show the child Plate 25, and say: All right, let's try this one. Point to "cone," and record the child's response verbatim. Administer all items, through number 60, following the procedure outlined above. Note: Plates 33, 35, 38, 44, 48, 54, and 56 are not shown to the subjects.

6. On items which have "probes" indicated, administer the probe for any response which cannot clearly be scored as 2. Ask the probe exactly as stated on the answer for. Where more than one probe is given, ask one or the other probe but not both. If a probe is asked, write "(Q)" preceding the subject's response.

PPVT-Expressive Test

Items and Scoring Criteria

CONE

2 - Cone plus ice cream.

1 - Cone or ice cream.

ENGINEER. Probe. "What kind of man"

2 - Engineer. Train man. Bus man.
Response relating man to vehicle in driving capacity, i.e., truck driver.

1 - Man and vehicle mentioned but driving function not specified, i.e., man in a train.
Describes man's role as function of uniform, i.e., milkman

PEEKING. Probe: What is she doing

2 - Girl plus activity, i.e., hiding, peeking.
Correct activity (hiding, peeking) and related to the tree.

1 - Girl in some other activity, i.e., climbing the tree, holding the tree. Boy plus hiding or peeking.

8. KITE. Probe: What does it do

2 - Kite plus activity (fly)

1 - Kite or activity.

9. RAT

2 - Rat, mouse (hampster, guinea pig)

1 - Other mammals similar in configuration, i.e., squirrel, rabbit, raccon.

0. TIME. Probe: What does it do or what is it for.

2 - Clock or Watch plus tells time.
Time as a sequential event.

1 - Clock or watch.

1. SAIL. Probe: "What kind of _____"

2 - Boat plus sail (or wind blown). Sailor boat.

1 - Boat. Ship.

32. AMBULANCE. Probe: "What kind of truck (or car)"

2 - Ambulance.
Car or truck plus function, or adjective describing medical function, i.e., car to take sick people to the hospital, doctor's truck, hospital car.

1 - Car or truck with emergency but non-medical function, i.e., firetruck, police car.

34. SKIING. Probe: "What is he doing"

2 - Boy (or girl) plus winter snow activity, i.e., skiing, sledding, sliding down the snow.
Skiing.

1 - Other winter activities, i.e., ice skating, sliding down the hill, playing in the snow, with or without mention of the boy.

36. TWEEZER. Probe: "What is it called"

2 - Tweezers. Any gripping tool, i.e., pliers.

1 - Function dexcribed without specific mention of the tool, i.e., gets splinters out, takes out teeth.

37. WASP.

2 - Response must specifically indicate a flying insect with a stinging function or mention one of these specific insects - wasp, bee, fly, hornet, mosquito.

1 - Flying insects, i.e., butterfly, moth
Insect or bug (spider or ant not acceptable).

39. PARACHUTE. Probe: "What is it called" or "What is it doing"

2 - Parachute plus activity, i.e., carrying, floating, coming down.

1 - Parachute or activity.

40. SADDLE. Probe: "What is it called."

2 - Saddle.
Description of function (horse and seat), i.e., horse thing to sit on.

1 - Response indicating relationship to horse, i.e., thing that goes on horse, cowboy sits on it and rides, horse thing.

41. TEMPERATURE. Probe: "What is it called"

2 - Thermometer.
Tells or measures temperature, heat or cold.

1 - Temperature.
Functional response which does not specifically mention thermometer or measurement of temperature or heat, i.e., tells you when to put your coat on.

CAPTAIN. Probe: "What kind of man"

2 - Captain.
Man related to unifcrm, ie., policeman, mailman.

1 - Man.

WHALE. Probe: "What kind"

2 - Whale. Porpoise.

1 - Fish. Flipper. Seal.

BALANCING. Probe: "What is he doing" or "What is he called"

2 - Response must mention seal and ball related in an effort activity, i.e., balancing, holding, bouncing, playing with the ball.

1 - Seal (or similar animal,i.e., walrus) and ball related but effort activity not explicit, i.e., seal with a ball.
Animal (unspecified or wrong animal) in activity with the ball, i.e., animal bouncing the ball on his nose, skunk playing with the ball, i.e., Correct activity related to the ball without mention of the animal, i.e., turning the ball around on his nose.

COBWEB. Probe. "What is it called"

2 - Cobweb. Spider web, net, or house. Web.

1 - Umbrella. Spider thing. Insect house or equivalent, i.e., bug lives here. (spider alone not acceptable).
Function described.

PLEDGING. Probe: "What is she doing"

2 - Girl plus activity toward the flag, i.e., girl saying prayers to the flag, girl pledging allegiance.
If the girl not mentioned, response must state "saying pledge to the flag" or "pledging allegiance".

1 - Girl plus activity not related to the flag, i.e., girl saying prayers. Girl and flag mentioned but unrelated.
Description of activity, i.e., putting her hand over her heart.

HYDRANT. Probe: "What is it called: or "What does it do"

2 - Hydrant (water pump or water thing) plus function, i.e., water tank to spray out fire, thing fireman puts hose to for water, water pipe to clean streets.

1 - Hydrant, water thing or object but limited to water source, i.e., for water to come out, that sprays water.
Description of function, i.e., for cleaning the streets, puts out fire, squirts water in the summer.

50. BINOCULAR. Probe: "What does it do"

2 - Binoculars. Response specifying magnification function.

1 - Glasses. Things to look through. Thing to see pictures in.

51. LOCOMOTIVE.

2 - Locomotive. Train. Engine.

1 - Choo-choo

52. HIVE. Probe: "What is it doing"

2 - Flying insects with box or hive in a related activity, i.e., live there, make honey, going home. (Response must indicate functional relationship between hive and bees.)

1 - Bees, and-or hive.

53. REEL. Probe: "What is it for" or "What is it called"

2 - Reel. Functional description of line as used in fishing, i.e., something to pull in the fish.

1 - Functional description of reel but not related to fishing. Response hiving to do with fishing rod. Pencil sharpener.

55. GNAWING. Probe: "What is he doing" or "What is he called"

2 - Beaver in activity with the tree, i.e., eating, cutting, chewing, chopping or getting tree down.

1 - Correct activity but animal is wrong or not specified, i.e., squirrel chewing down the tree, animal eating the tree, gnawing the tree. Beaver mentioned but activity incorrect or not related to the tree (beaver jumping in the tree)

57. BANNISTER.

2 - Bannister.
Steps plus rail or pole or functional description of rail, i.e., thing you hold on to going up stairs.

1 - Steps or stairs.

58. IDOL. Probe: "What kind of man (or lady)

2 - Idol. Statue. Non-living representation of a person.
1 - Indian. Genie. King, Queen, (God or Jesus, angel)

59. GLOBE.

2 - Globe. Round world or map
1 - Ball-map. Earth. Moon. World. Map.
Functional description, i.e., thing that goes around and tells you where you are.

60. WALRUS

2 - Walrus. Sealion. Seal 1 - Flipper. Porpoise.

TABLE I

PPVT-Expressive

Reliability Coefficients for Pre & Post Test Scores

	Split-half	Kuder-Richardson
Pre-test	r=.78	.88
Post-test	.77	.89

TABLE II MEAN STANFORD-BINET IQ's by TESTER

Pre-test

Tester	N	Mean	Standard Deviation
02	13	84.08	9.51
04	68	91.13	10.50
05	54	94.52	15.64
06	12	93.25	7.80
08	5	77.80	3.12
09	39	86.82	15.20
11	16	87.75	9.58
12	8	77.88	10.87

Post-test

01	7	102	12.60
02	56	89.27	12.67
03	6	85.17	9.19
04	17	94.88	13.69
05	43	91.93	15.71
06	17	94.00	9.91
07	8	88.25	7.79
10	9	90.56	10.80

TABLE III MEAN PPVT-R IQ's by TESTER

Pre-test

Tester	N	Mean	Standard Deviation
20	36	82.47	12.70
21	15	93.87	9.76
22	101	83.60	16.69
25	8	81.63	13.07
26	31	85.71	18.51
27	16	91.00	11.85
30	4	92.00	4.58

Post-test

23	28	83.75	18.19
24	63	94.92	13.42
28	47	95.47	15.16
29	19	89.42	10.85
31	27	83.70	15.49

TABLE IV MEAN PPVT-E RAW SCORES by TESTER

Pre-test

Tester	N	Mean	Standard Deviation
41	27	26.59	5.44
42	35	27.51	6.91
43	6	22.67	1.60
44	85	30.12	5.07
45	50	26.78	5.33
46	15	17.73	4.84

Post-test

40	49	28.31	5.24
41	14	24.14	3.42
43	30	29.83	5.73
44	65	33.28	6.06
45	6	35.67	7.27

TABLE V

Behavior Inventory

Factor Structure

Factor I: Withdrawing, Inhibited, Casper Milquetoast
(Shared Variance=34.67)

Item	
22. Is constricted, inhibited, or timid; needs to be urged before engaging in activities.	-.826
24. Is reluctant to talk to adults; responds verbally only when urged.	-.811
7. Often keeps aloof from others because he is uninterested suspicious, or bashful.	-.772
49. Approaches new tasks timidly and without assurance; shrinks from trying new things.	-.764
46. Is lethargic or apathetic; has little energy or drive.	-.728
35. Is eager to inform other children of the experiences he has had.	+.684
39. Asks many questions for information about things, persons, etc. (Emphasis here should be on questions prompted by genuine curiosity rather than bids for attention.	+.678
5. Talks eagerly to adults about his own experiences and what he thinks.	+.678
34. Often will not engage in activities unless strongly encouraged.	-.677
1. Is usually carefree; rarely becomes frightened or apprehensive.	+.595
33. Likes to talk with or socialize with the teacher.	+.576
28. When faced with a difficult task, he either does not attempt it or gives up very quickly.	-.565

Factor II: Hostile, Impulsive, Low controlled, Donald Duck (Shared Variance=10.52)	
Item	
36. Emotional response is customarily very strong; over-responds to usual classroom problems, frustrations, and difficulties.	+.686
19. Is excessive in seeking the attention of adults.	+.647
18. Responds to frustration or disappointment by becoming aggressive or enraged.	+636
16. Has little respect for the rights of other children; refuses to wait his turn, usurps toys other children are playing with, etc.	+.615
26. Is often quarrelsome with classmates for minor reasons.	+.601
10. Is jealous; quick to notice and react negatively to kindness and attention bestowed upon other children.	+.593
37. Is uncooperative in group activities	+.570
23. Is even-tempered, imperturbable; is rarely annoyed or cross.	-.568
Factor III: Non-persistence in Problem Solving (Shared variance=4.98)	
11. Is methodical and careful in the tasks that he undertakes.	+.708
20. Sticks with a job until it is finished.	+.642
40. Usually does what adults ask him to do.	+.631
45. Is wanted as a playmate by other children.	+.631
13. Tries to figure out things for himself before asking adults or other children for help.	+.586
2. Is sympathetic, considerate, and thoughtful toward others.	+.576
25. Works earnestly at his classwork or play; does not take it lightly.	+.562

TABLE VI

Stanford-Binet FACE SHEET

FACTOR STRUCTURE

	Loading
Factor I. Emotional.	
Emotional independence; self confidence; needs little ego support for problem solving.	
8. Assured	.840
6. Realistically self-confident	.731
7. Comfortable in adult company	.623
13. Needs minimum of commendation	.567
Factor II. Problem Solving.	
Problem solving persistence; response to challenge of problem; task oriented (intermediate risk taker).	
11. Eager to continue	.797
12. Challenged by hard tasks	.734
9. Persistent	.701
10. Reacts to failure realistically	.576
Factor III. Social.	
Initiates activity; quick to respond; socially confident.	
3. Initiates activity	.783
5. Socially confident	.642
4. Quick to respond	.640
Factor IV. Attention.	
Absorbed by task; normal activity level; attention specificity.	
2. Normal activity level	.693
1. Absorbed by task	.686

APPENDIX

Younger Sample:

A total of 79 children of the 90 who were eligible were found and tested. The purpose in selecting a population who experienced a summer Head Start program, who did not attend kindergarten or another pre-school program the following year, and who might return the subsequent summer for a second Head Start experience before entering kindergarten, was to study the effect of multiple summer programs. This population represents the first stage of a longtitudinal study and is reported here in the form of a progress report. Table VII (Appendix) summarizes the several measures taken of these children in the late fall after their first Head Start experience.

Of interest here is the rough comparability of several of the measures between this group and the older childrer. Means for the several groups are the same, and all differences between boys and girls in the younger group are in the same direction as in the older group. The discrepancy between the Binet and PPVT-R in the high and low Binet younger children corresponds to this discrepancy in the older children. The only inconsistency is in the PPVT-E and PPVT-R discrepancy. The younger population has no discrepancy between these measures above or below the median Binet IQ. It should be recalled that for the older children, those below the median Binet were superior on the E, whereas those above the median Binet were superior on the R. These differences may reflect the age differences in the populations and will be examined further as subsequent data become available.

TABLE VII. MEANS, STANDARD DEVIATION AND DISCREPANCY SCORES PPVT-R, PPVT-E, STANFORD BINET (Younger Sample)

	N	PPVT-R	N	PPVT-E	N	Stanford Binet	SB-R Discrep.*		E-R Discrep.*	
BOYS	54	84.50 20.1	42	22.50 5.87	45	87.24 13.3	(Median Binet IQ = 87.8)			
GIRLS	26	80.42 16.9	31	19.90 6.53	34	89.12 16.8	Above Binet Median	Below Binet Median	Above Binet Median	Below Binet Median
TOTAL	60	82.73 18.9	73	21.40 6.27	79	88.05 14.93	+7.24	+2.03	-0.21	+0.19