DOCUMENT RESUME

ED 037 239 PS 002 825

AUTHOR Datta, Lois-ellin
TITLE A Report on Evaluation Studies of Project Head Start.
SPONS AGENCY Office of Economic Opportunity, Washington, D.C.
PUB DATE 69
NOTE 26p.; Paper presented at the 1969 American Psychological Association Convention, San Francisco, California

EDRS PRICE EDRS Price MF-$0.25 HC-$1.40
DESCRIPTORS *Compensatory Education Programs, *Federal Programs, Followup Studies, Intervention, Longitudinal Studies, *Preschool Programs, *Program Evaluation, Research Needs, Research Problems, *Research Reviews (Publications)
IDENTIFIERS *Head Start

ABSTRACT

Evaluation of Head Start has been based on four sources of information: (1) census surveys of children and families served and programs offered, (2) special research, projects on child development and experimental programs, (3) a longitudinal study of the development of low income children, and (4) a series of national evaluation studies. Available data appear to indicate that Head Start and other preschool programs have an immediate impact, but little is known about why, or under what circumstances optimum results may be obtained. Sustained gains are still being sought. Children who have not attended preschool programs tend to catch up in primary school with attenders, but little is known about why this happens. A planned variation study is in progress comparing children in sponsored Head Start and sponsored Follow-Through classes and children attending "regular" Head Start and "regular" primary schools. Head Start evaluations have tried to locate program variations other than administrative which may affect child development. Considering the evidence now available, the assumptions on which Head Start was based still seem tenable. Research is needed to clarify relationships between program and child variations, and the effects of long-term interventions. (NH)

A Report on Evaluation Studies of Project Head Start

Project HEAD START
Office of Child Development
U.S. Department of
Health, Education, and Welfare
Washington, D.C. 20201

A REPORT ON EVALUATION STUDIES OF PROJECT HEAD START*

Lois-ellin Datta**
National Coordinator, Head Start Evaluation

The past ten years have seen the rise of programs seeking to make a significant difference in the lives of the poor. Prominent among these are attempts to accelerate the cognitive development and scholastic achievement of children from low-income families. Not all of these programs are in the narrow sense "compensatory" in philosophy or approach, although they have been discussed under this label in assessments of the effectiveness of current strategies.

We have been told recently that compensatory education has been tried and that it apparently has failed; that Head Start as an example of compensatory education is ineffective; and we have been encouraged to seek new strategies. Jensen (1968) recommends training to foster special skills for different ethnic groups. Jencks (1969) urges that we look away from the schools to other scenes, particularly the family and the neighborhood. Still others direct attention to maternal nutrition and the well-born child, and to parent training in infant education.

Many of these are areas to which attention has been overdue. There are those who feel, however, that we are in danger of being too hasty in writing off compensatory education and in turning away from efforts to understand what may be the most effective preschool experiences both immediately and in the long run. These arguments are based in part on the assumption that although Head Start may have been oversold or may not be the success that we hoped, some compensatory education programs are at least a fair success (Hunt, 1969). Still other reviewers judge that the data are not all in, or not in enough to justify epitaphs on compensatory education. In Kagan's words (1969), "The value of Head Start or similar remedial programs has not yet been adequately assessed."

---

*Paper presented at the 1969 American Psychological Association Convention.

**The Head Start approach to evaluation has been shaped by many researchers; particularly influential in setting the course outlined in this paper were Dr. Edmund Gordon, Dr. Edward Zigler, Dr. Urie Bronfenbrenner, and Dr. John McDavid, Director of Head Start Research and Evaluation from 1967 to 1968. I also wish to acknowledge the contributions of Dr. Edward Suchman, Dr. Boyd McCandless and Dr. Alfred Yankauer of the Head Start Research Advisory Council and the Directors of the Head Start Evaluation and Research Centers; Dr. Herbert Zimiles, Dr. Frank Garfunkel, Dr. Carolyn Stern, Dr. Dorothy Adkins, Dr. Russell Tyler, Dr. Robert Boger, Dr. Myles Friedman, Dr. Edward Johnson, Dr. William Meyer, Dr. Theron Alexander, Dr. John Pierce-Jones, Dr. Shuell Jones, Dr. Robert L. Thorndike and Dr. Virginia Shipman.

What Head Start has undertaken in the way of assessment appears not to be widely known. Perhaps a discussion of the Head Start evaluation and research effort and of the findings as we see them will be of value to the broader discussion.

Head Start has been in operation for five summers and four full years. It would seem reasonable that there should be at this time reliable evidence on the immediate and long-range effect of Head Start as an approach. Balanced against this reasonable expectation are some unreasonable realities. Some are well-known to any researcher; others may be less obvious. First are the formidable problems in organizing and administering a nation-wide preschool, community-controlled, comprehensive program. As one example of the way in which these matters can affect evaluations, consider the implications of funding uncertainties on local program operations. Funding delays reduce the lead time for recruiting and staff training, and in some instances, actual length and stability of operation.[1] Second, we are learning our way in training community people for positions in the classroom and in program administration. Third, the field of education for preschool disadvantaged children has been created almost from the ground up in terms of available courses and qualified training staff. Fourth, our measures of product and process began from virtually nil and have developed only haltingly. And fifth, many studies of necessity are compromises between designs required for statistical inference, Head Start's outreach to eligible children[2] and community control of program decisions.

In a very real sense, Head Start as an approach has not been tried. Four years is a brief interlude in which to create a new field, to develop new careers for thousands of poor, to explore ways to maintain program quality for Head Start in the inner cities of our country and in communities so remote that even the mails don't always get through. That Head Start exists at all has been described as a minor miracle; that it is developing toward the carefully planned well-supported operation it was intended to be does honor to the many hours professionals, paraprofessionals, and parents have given to Head Start. That Head Start and the field of preschool evaluation both have some miles to go before they are ready to keep the promise of a national assessment of Head Start as the exemplar of preschool education should seem obvious to anyone who has ever tried to implement a program on even a small scale and to anyone who has struggled with measurement. As Edward and Mary McDill and Timothy Sprehe (1969) write:

> . . . compensatory educational programs have been put in a position never demanded of educators before. No public school system has ever before been abolished because it could not teach children to read and write. Yet compensatory programs, aimed at the very children who are going to be losers in the regular school program, are in just this situation. The programs are being asked to succeed in a shorter time than that which the regular school systems have had. Perhaps this is healthy. Insisting on nothing less than success as a

> condition of survival is indeed a great motivator for achieving success. But outright condemnation of all compensatory programs should be tempered by the realization of the magnitude of the task with which they are confronted and the short time they have been coping with the task. (pp. 38-39)

Rather than belabor at this time the obvious and hidden hazards to evaluation, I'd like to review the studies Head Start has undertaken and what our interpretations of the available data have suggested.

Head Start research and evaluation has had four major components: surveys, research, a longitudinal study, and national evaluations.

(1) Census Surveys. A series of descriptive studies of a nationally representative sample of Centers has been conducted for Head Start by the Bureau of Census. Selection of Centers, questionnaire distribution, follow-up and analyses are handled by Census procedures. The content of the questionnaires is provided by Head Start program specialists. The questionnaires primarily obtain information on compliance with Head Start guidelines with regard to the children and families served and the programs in the major areas: health, nutrition, volunteers, parent participation, social services and education. The surveys have been conducted for ten program periods, every summer and full-year since 1965. A report by Barbara Bates of our office in cross-tabulated detail for summer and full-year, part-day and full-day programs is now available through the ERIC system (Bates, 1969).

(2) Research. Head Start has supported research studies on child development, on instrument development, pilot projects, demonstration projects and most recently, transitional studies designed to explore how to minimize dilution of program quality when programs move from the laboratory to the field. Copies of reports of all completed projects are available through ERIC.

(3) Educational Testing Service Longitudinal Study. The third major effort is a longitudinal study of the development of low-income children, a project almost three years in preparation as a cooperative effort between the Head Start Research Advisory Council and the researchers at Educational Testing Service. The study will follow all children in a target area from the first observations at age 3 1/2 through their school experiences to the end of the third grade. The project may contribute to instrument development and to our knowledge of child development; it will explore the associates of different preschool and school paths the children can take. In each of the four target areas, about 50% of the children are expected to attend Head Start; the Follow-Through program is also available to about 50% of the children in three of the communities. A two-volume ETS report on the design, on the conceptual approach that has shaped the selection of measures in each of the domains, and on the analytic model is available through ERIC.

(4) National Evaluation Studies. The fourth major area is national evaluative research studies. These studies began with the first 1965 effort to assess the average change associated with summer Head Start. We are wiser now and proceed on the basis of the following three assumptions:

a. that Head Start programs are diverse in their specific objectives and thus in experiences provided to the children,

b. that even where goals are similar, success in implementation may vary, and

c. that children and their families are diverse in ways likely to interact with the effectiveness of any single, well-implemented approach.

This awareness has shaped the change in Head Start efforts from summative evaluations to evaluations directed to the question, "What in the diverse program approaches makes what kind of difference in the ways in which children and their families may change." The most recent national studies are designed to describe what is happening to the children and to relate differences in what is happening to differences in outcome.

Details of the 1967-68 Study. The 1967-68 evaluation study began with collection of data on the teachers, physical sites, children and programs of candidate centers and classes. These data were provided to the 14 Head Start Research and Evaluation Center Directors by the Head Start regional staff and by their own information networks, and were reviewed by the E&R Directors, by the Director of Head Start Research and Evaluation and by members of the Head Start Research Advisory Council. Classes varying as widely as possible in anticipated educational approach and child characteristics were selected as sample classes.

Criterion Measures. Each of the 14 University-based Evaluation and Research Centers collected pre and post a common core of data on about 150 children. The measures were the Stanford-Binet, a rating of child behavior in the testing situation, and the Social Interaction Observation Protocol. The SIOP which was developed at the University of Kansas records the rate and content of peer and adult social initiations and responses for a 45-minute free-play observation period for each target child. The identity of participants in the interactions is noted so a detailed sociometric also can be constructed. This information is costly and difficult to collect but provides a direct observational record of social responsiveness. The common core data also included an initial and final interview with the child's mother (demographic data, the Hess-Shipman educational attitude scales, their "First Day" question and the Sigel child-rearing practices items).

Process Measures. As common core process measures, all Centers collected five days of observational data with the Observation of Substantive Curricular Input (OSCI) form developed by Dr. Carolyn Stern of UCLA and

a committee of E&R Directors. The OSCI uses the observer as a camera to scan class activities. There were 35 three-minute observation periods for each day, and five observation days per class throughout the year for a total of 175 three-minute segments of each of the sample classes. The observer began each three-minute scan with the largest group in the program, recording group size, context of the group's activity, content of the activity, whether it was teacher or child controlled and the materials involved. The scans continued with the next largest group, and the next and so on for the three-minute period. Each three-minute scan could potentially yield from one record (whole group activity) to 15 records (each child doing his own thing). During the past year, Dr. Stern has been editing the OSCI records, assessing the reliability of each of the 300 cross-code combinations within observers, within days, within classes, within Centers and so on, and then has worked toward combining the best of these codes to identify experiential clusters of Head Start classes. All Centers also collected demographic data on the teachers, an inventory of the class and Center physical facilities and information on the individual children and their participation.

Other Measures. While this battery placed heavy demands on the E&R Centers, it was still felt that the information in any single area was too shallow. To enrich measurement without overloading the children or their own staff, the E&R Centers formed five clusters. Each Center in a cluster collected additional common data on about half of its evaluation sample. The Curriculum I cluster collected Dr. Frank Garfunkel's Classroom Behavior Survey which describes critical teacher/child and child/child interactions on a variety of dimensions. The Curriculum II cluster obtained individual child OSCI's which will permit methodological studies based on the individual's experience as contrasted with predictions based on global descriptions of the class, and the Observer's Rating Form developed at the University of Texas. The Social-Emotional cluster collected mother-child interaction data on the three Hess-Shipman tasks: toy-sort, block-sort, and Etch-a-Sketch; the Brown IDS self-concept measure; the Picture Playboard Sociometric developed at Michigan State University; and a mid-year SIOP. The Cognitive I cluster obtained data on the Sigel Picture Categorization test, the Pictorial Test of Intelligence, the Animal House, Picture Completion, Mazes, Geometric Designs, Block Designs and Sentences WPPSI subtests; the Auditory-Vocal and Visual-Motor Sequencing subtests of the ITPA and, from the Leiter International, subtests III-3, IV-4, V-1 and V-3. The Cognitive II cluster obtained an abbreviated PSI, the ITPA motor-encoding and vocal-encoding tasks, the Maccoby-Moss Draw-a-Line-Slowly measure, the Information, Animal House, Mazes, Geometric Design, and Block Design WPPSI subtests, the Draw-a-Man test and, from the Leiter, subtasks IV-1, IV-2, IV-4, V-2 and V-4.

While this design was a quantum change from earlier studies, the 1967-68 field experience demonstrated the difficulty in a naturalistic design in predicting the actual content of the classes and in obtaining enough variation where it was needed to avoid confounding child, regional, and program characteristics. It became clear indeed that one began by designing a

study permitting comparisons among different approaches, and that such a strategy made far better sense for Head Start than reliance on covariational or regression techniques.

Design of the 1968-69 Evaluation. The 1968-69 national evaluation thus represented another quantum change. Each of the E&R Directors either identified a reasonably assured natural variation or proposed a direct intervention. Each developed a research design appropriate for his study that could stand on its own as an investigation of "what works best." The studies were linked as a national effort in four ways: first, by common and extensive measures on the children, their families and the classroom experiences; second, by inclusion of non-intervention "regular" Head Start comparison classes in each sample; third, by the common pre/post design involving comparisons among distinctive Head Start programs; and fourth, by common assessment of the extent to which each Center's variation was implemented in its own classes and was occurring "spontaneously" in every other sample class. For example, one aspect of the Tulane-South Carolina-Texas cooperative study involved a motivation training program. What should have happened in the motivation training program classes was stated operationally by the researchers. The post program teacher interview and observer rating forms collected five times over the year both included items based on these statements, with data collected for all sample classes. These data will eventually permit comparison of three groups:

1. classes homogeneous for variation created by the researcher,

2. classes homogeneous for similar events occurring without researcher intervention, and

3. classes in which there is no evidence that such variation occurred, including those in which the researcher attempted but failed to make something happen.

A given sample class may be a researcher variation class in one analysis, a natural variation class in another analysis, or a comparison class in a third analysis. In addition to the common core measures, each E&R Center collected data of criterion relevance to its own study.

The individual Center reports for the 1968-69 study should be available in early 1970 through ERIC. The national analyses are to be undertaken centrally and will involve meshing findings from interactive analyses of the 1966-67, 1967-68 and 1968-69 studies. It is likely to be some time before this report is available: analysis with new measures is a slow process, and the systems analysis relating classroom experiences to change is undoubtedly going to be an arduous task to do well.

Programs can not wait for evaluative research findings, however, even if the better study seems beyond the other mountain. Nor, in fact, have they. The first reports of the follow-up studies of summer 1965 Head Starts

stimulated the development of the Full-Year program and of Project Follow-Through. Other studies provided the impetus for the experimental Parent-Child Centers, extending Head Start downward to families with infants from 0 to 3 years of age. For some months now, we have been reviewing preliminary national data on Head Start classes and assessing the implications of other studies of Head Start and preschool intervention. Three questions of particular interest to us related to the variability of Head Start classes, to the immediate impact of the program, and to the children's performance in primary school.

Are Head Starts Heterogeneous? Sigel has noted, "The learner, the program and the teacher function in an educational setting which has its own institutional arrangements. Systems vary in the degree to which they are open to change and willing to modify the curriculum, willingness to re-orient resources and change priorities. Teachers in these systems vary in the degree of independence as well as skill and morale. In other words, we have in the educational system of things tremendous hetero-geneity. All of these factors contribute to the degree to which new and innovative programs can be successful. Thus compensatory educational programs vary from community to community as well as within communities. It is difficult, if not impossible, to expect uniform gains. . . . Thus, one reason why we can't make . . . generalizations is because of the heterogeneity of samples and environments." (1968, pp. 17-18.)

In practice, the assumption of diversity requires some test lest it becomes an excuse for failing to confront facts. The data I am about to describe were collected on 260 classes included in the 1967-68 E&R evaluation.

- The median teacher was between 28 and 33 years of age. Some (3%) were less than 21 while as many as 19% were over 60. The majority (55%) were white; 40% were Negro. Most received a B.A. degree (67%) but there was considerable regional variation. In classes studied by seven E&R Centers, from 43% to 87% of the teachers had only completed high school while in classes sampled by four other E&R Centers, 75% or more had Bachelor's degrees.

- The majority of teachers had had one or more full years of paid experience with children. More than half had been employed with Head Start for more than a year (64%); the range across E&R Centers varied from 32% for one area to 79% for another.

- Attrition varied from 4% for one area to 19% for another. While 80% of the total sample of children were reported to have attended 4/5 or 5/5 days a week on the average, reported absenteeism ranged from over 60% for some sites to 0% for others. Class stability also varied: in classes in a large Northern city, for the majority of the 127 sample children over 50% of their classmates at the end of the year were different from their classmates at the beginning of the program. In classes studied in one geographically isolated area, 100% of the 136 children had 85% or more of the same class-mates throughout the year.

- Class structures represent another potentially significant area of diversity. Sixty-two percent of the children attended ethnically homogeneous classes (defined as 75% or more of the children from one ethnic group); 38% attended ethnically mixed classes. If one wished to study this variable in the 1967-68 sample, 18% of the children were white children in a majority white class with a white teacher while 24% were other children in a majority other class with an other teacher; 19% were other children in a majority other class with a white teacher; and 12% were other children in a mixed class with an other teacher.

- Duration of class might be another variable of relevance to child development: 58% of the children attended classes which met from three to four hours daily; 16% attended classes meeting from five to six hours daily; and 15% attended programs which met from seven to eight hours.

The following data come from the 1967-68 Census survey for children who varied in ethnicity (24% White, 51% Negro, 10% Mexican-American, 6% Puerto Rican, and 7% American Indian children), experience (of all children, 18% had previous Head Start experience prior to the sample year, 20% had previous nursery or day-care experience and 60% had neither); and family pattern (30% of the children came from mother-only homes, 60% from nuclear family homes and 10% from homes with extended families).

- Paternal education varied from less than sixth grade through some college (7%). The median family size was six persons: the range was from two persons (2%) to more than 13 (4%). About 30% of the mothers were employed and about 60% were housewives. Many children came from families whose siblings had previous Head Start (23%) or other preschool experience (27%). Only 49% of the 1967-68 Census sample came from families with no previous Head Start or Day Care participation. About 27% were only children but virtually all of the others shared parental attention with one or more siblings under six years of age. Some index of physical status may be reflected in the fact that 45% of the mothers reported that something wrong physically with the child had been identified on the Head Start physical examination.

- According to the 1967-68 Census sample, when Center Directors were asked to check as many labels as would apply to their programs, about 9% checked Montessori, about 13% group day-care, about 40% responsive environment, about 15% structured drills and 61% environmental enrichment. With regard to curriculum emphasis, the majority of all Directors reported attempting to influence sensory-motor development, language development, social skills, concept development, self-esteem and motivation, while only 50% indicated that the development of pre-academic or academic skills was an important goal.

Preliminary findings from the 1967-68 OSCI records for 136 E&R sample classes indicate that the apparent similarity among emphases as described by program directors does not reflect the diversity of program input experienced by Head Start children as seen by observers in the classroom. The basic OSCI distribution is percent of the total record units in which a given activity was observed. Since each record unit could include one or two activities, the percents will exceed 100%; this is presumably appropriate to the fact that a given activity may have more than one salient component. For example, a child and an adult are at the water-play table. If the teacher directs the child's attention to the properties of wet sand, labelling these properties and eliciting verbal responses, the activity could be coded as small muscle development and as informal language development. If the teacher says nothing during the scan, the activity would be coded only a small muscle development. While the OSCI is a complex measure whose potentials and pitfalls are not yet fully explicated, some highlights may suggest something of the observed programs.

- Caretaking was a low frequency activity with less than 5% of the activities falling into categories such as arriving, clean-up or toileting. Primarily undifferentiated activity such as fighting occurred in less than 7% of the scans.

- Many activities occurred with moderate frequency and showed considerable variation. For example, the modal time spent in dramatic role playing was 15-20% (20% of the classes); however, 6% of the classes had virtually no incidents of dramatic role-playing while another 7% had dramatic play observed between 35-40% of the time.

- Very few classes were observed to spend more than 5% of the time in specific training for auditory discrimination, quantitative development and scientific activities; however, as many as 20% of the classes would form a cluster in which these directed kinds of training were of relatively high frequency. Visual perception, on the other hand, varied from less than 5% of the activities (3% of the classes) to 30-40% of the activities (4% of the classes).

- The most widely dispersed activities were motor, rote, informal verbal development and social interactions. The amount of language training in the formal sense varied from less than 5% of the activities to between 25 and 35% of the activities (6% of the sample) with the mode at between 10 and 15% of the activities. Informal language development was an almost rectilinear distribution ranging from 5% to 75%; some Head Start sample classes apparently had teachers who used virtually every opportunity to facilitate language development while other teachers made virtually no attempt to use the opportunities in this way.

- Emphasis on good conduct (rules and regulations) varied from little or none (12% of the classes) to as many as 30% of the incidents (9% of the classes) in a positively skewed distribution. In no class was the locus of control always observed to be the child; this distribution was symmetric and bell-shaped, with the median at 50% of the incidents being teacher controlled. Some classes would appear to be substantially teacher controlled while others could be meaningfully classified as very low on teacher control.

- Group size is still another variable of potential educational significance. The number of activities tallied as "whole group" varied from less than 5% (in three classes) to between 65% and 70% (in two classes); the distribution on this code is flat and somewhat positively skewed.

Available data on these relatively crude structural measures are consistent with the assumption that Head Start programs have varied in ways considered to be educationally significant. If such factors as parent participation, teacher skill and control techniques were added, it is likely that the diversity would be still greater.

Review of the Immediate and Long-range Impact of Head Start. Miller (1968) has noted:

> In our work with various groups of children from disadvantaged environments, we have found that it is not much of a trick to obtain an average Binet I.Q. score gain of 15 to 20 points over a year intervention. This is consistent with other findings and appears to be about the asymptote which is generally obtained. The real trick is to maintain these gains over a period of time so that the usual picture of progressive decline does not emerge. . . . (p. 17)

Miller was reviewing findings of such experimental programs as Susan Gray's at DARCEE and David Weikart's studies at Ypsilanti. We are less certain about what is and is not an easy trick with regard to Head Start. While our uncertainty will perhaps be considerably reduced on completion of such studies as the national analyses, a summary of what available evaluations of Head Start appear to show about immediate impact and retention of gains seems appropriate.

The Immediate Impact of Some Head Start Programs*

- Many though not all studies of summer Head Start programs show that children's performance on general ability tests improved significantly, although the scores typically did not reach the national averages (Chesteen, 1966; Eisenberg, et al., 1966; Hodes, 1966; Berlin, 1965; Horowitz & Rosenfeld, 1966; Cawley, 1966; Berger, 1965; Harding, 1966; Pierce-Jones et al., 1966; Temp & Anderson, 1967).

- Jensen and Kohlberg (1966), Beller (1967), Bittner & Rockwell (1968) and Nalbandian (1968) have reported a smiliar pattern for Full-Year Head Start programs; Alexander (1968), Faust (1968), studying 1967-68 programs, found that after Full-Year Head Start, the children's performance reached the national average on the Stanford-Binet (IQ 100). In these reports, there is a common element of reliable gains for both summer and the full-year programs; there is also some indication that the final level of achievement is a function of the length of time in the program, at least in the six weeks to nine months' range represented.

- Some additional support for this interpretation is found in preliminary data from the national studies. These analyses indicate that children without previous nursery or Head Start experience had average IQ scores of about 86 when tested in the first two weeks of the program while children who were tested for the second time after about 40 weeks had average IQ scores of about 103. The cross-sectional curves (excluding drop-outs) are significantly linear with some indication of a plateau after about 24 weeks and an acceleration after about 36 weeks.

- In the areas of attitudes, motivation and social behavior, there is some evidence that Head Start was associated with immediately apparent changes. The primary source of this evidence is teacher ratings of the children (Berlin, 1965; Harding, 1966) since other measures have proved to be unreliable (Harding, 1966; Hess, 1966; Chorost, Goldstein and Silberstein, 1967). The children were reported to show more socially appropriate behavior following

---

*The following sections owe much to Dr. Edith Grotberg's (1969) summary of the findings of studies funded by Head Start and to unpublished reports by Richard Armstrong. I am grateful to both Dr. Grotberg and Mr. Armstrong for permission to quote without direct attribution from their work, and to Mr. Armstrong for extended discussions of methodological issues in evaluation designs.

their experience in Head Start, including increased interest in new things (Harding, 1966; Soule, 1965), improved adult-child and child-child interaction patterns (Harding, 1966; Faust, 1968), increased task orientation (Horowitz & Rosenfeld, 1966; Ozer, 1965); improved attitude toward learning (Beller, 1968); and improved self-concept, decreased alienation from authority and increased trust in others (Lamb, Ziller & Maloney, 1965). Jensen and Kohlberg (1966) reported decreased task orientation but increased social interaction with the tester.

A majority of studies of Head Start have reported an immediate impact; data from the most recent studies of Full-Year programs indicate that performance tested immediately or soon after Head Start reaches the national averages on tests of general ability and learning readiness. These findings should not, however, be interpreted to mean that Head Start is a success or even that these particular Head Start programs are immediate impact successes. The reasons for this caution center around problems of design, in particular the impossibility in many instances of selecting at random some eligible applicant children to enter Head Start and others to be non-participant controls, and in other instances, the absence of either controls or satisfactory norms.

There are in addition to design considerations at least four alternative explanations of the reported immediate gains: (1) the difference between initial and final scores of the Head Start children, and between Head Start and comparison children, where available, represent changes in cognitive development and emotional maturity that are primarily attributable to the Head Start program; (2) changes occur but they are attributable to the new institutional experience and any such new experience, including much cheaper ones or the kindergarten or first grade all children will enter would do just as well; (3) Head Start children have become familiar with materials similar to those they encounter on the post tests and these specific skills rather than changes in overall development are being measured; and (4) there are powerful motivational factors associated with test performance for low-income children. Because psychological evaluations of preschool children can not be conducted impersonally with paper and pencil tests, the testing situation itself involves social interaction and the rise in scores may also reflect the increasing comfort the disadvantaged child feels with an often middle-class adult. It may also be that disadvantaged children are initially less motivated to perform well on tests but that Head Start experience enables them to become more task-oriented and more responsive to both the tester and the materials (Zigler and Butterfield, 1968).

Any one of these explanations is tenable and it may be that each contributed in varying degrees to the pre/post or Head Start/comparison differences in tests and observation scores. The magnitude of the difference in test/retest scores and the frequently reported failure of repeated testing of comparison children to be associated with performance change leads one

to question the test-retest explanation as the sole source of the difference.[3] None of these four explanations has yet been tested, however, with the experimental design necessary for unequivocal inference or with the measures that would permit separation of affective and more cognitive elements in performance. Nor do we know as yet what may facilitate the greatest changes: research offers some suggestions for further investigations particularly the value of preplanning and parent involvement (Grotberg, 1969), and these, together with the E&R studies may advance our knowledge still further. Whatever the explanation, our reading of available data is that the Head Start child is often likely to enter school with a greater cognitive and social readiness for learning, a readiness that may for recent Full-Year Head Start programs reach or exceed national averages on general measures.

The Longer-range Impact of Some Head Start Programs. While the evidence suggests some immediate changes in children attending Head Start programs, it has been typical since the first follow-up studies of the 1965 summer programs to find that this acceleration in rate of development was not sustained when the children entered primary school. What appears to happen is that the rate slows down for the Head Start children while their non-Head Start counterparts sooner or later catch up. While there are important exceptions to this finding (Beller, 1968), the majority of studies show that the developmental gap between Head Start and non-Head Start children is being closed or has been eliminated by the end of the first year in school, be it kindergarten or first grade (Wolff & Stein, 1966; Hess, 1966; Allerhand, 1967; Eisenberg, 1966; Hodes, 1966; Holmes & Holmes, 1965; Krider & Petsche, 1967; Morris & Morris, 1966; Jenson & Kohlberg, 1966; Chorost, Goldstein and Silberstein, 1967; Pierce-Jones, et al., 1966; Waller & Connors, 1968; Cline & Dickey, 1968; Sigel & McBane, 1966; Steglich, Cartwright & Allen, 1967; Cawley, et al., 1968; Coleman, et al., 1966; Bittner & Rockwell, 1968; Chesteen, 1966; Hubbard, 1967; Muse, 1968).

A number of explanations have been suggested for this "levelling off" phenomenon found in Head Start follow-up studies and for similar findings from many experimental preschool programs (Miller, 1969; Sprigle, et al., 1969; Gray & Klaus, 1969; Hodges, Spiker & McCandless, 1967; Karnes, 1967; Nimnicht, et al., 1967; Di Lorenzo, et al., 1967). The alternative explanations have included:

(1) One-time Impact. It has been suggested that changes which children experience in the preschool program would have occurred in kindergarten or first grade whether or not they had Head Start. A new environment, according to this interpretation, has a one-time, any-time impact.

(2) Class Norms. Since the teacher is primarily interested in the progress of the whole class, she must set the level of class activities below that necessary to challenge the more advanced Head Start children and give more attention to the group of children who are less advanced. Some evidence in support of this hypothesis is the finding (Wolff & Stein, 1967) that when 50% or more of the class had attended Head Start, the rate of gains was maintained, while when 25% or less of the class had attended Head Start, the differences were most likely to disappear.

(3) Peer Group Influence. The presence of more advanced Head Start children in a classroom may stimulate the development of non-Head Start children. Conversely, it is also possible that the Head Start children who can do many things feel less competitive pressure from their disadvantaged peers to develop new skills and abilities.

(4) Learning Cycles. If learning occurs in spurts followed by periods of consolidation, then during the first year of school, Head Start and non-Head Start children are at different stages of the learning cycle. With time, the development of Head Start children might again accelerate. Data from longitudinal studies (Beller, 1969; Sigel, 1967) tend to support this hypothesis; other, cross-sectional, data do not (Cicarelli, 1969).

(5) Factors in the School System. It may be naive to expect a child to continue to progress rapidly in a classroom where the teacher may be responsible for 30 or more children, may be primarily concerned with maintaining order and perhaps convinced that most of her students have little potential; and the demanding, active and inquisitive Head Start children may suffer more in this situation than non-Head Start children (Hyman & Kliman, 1967). A less extreme version of this interpretation is that the low-income child and his family require a different kind of program than that typically found in the school. It may be that when the child is provided over a period of time with the necessary attention from teachers who are adequately trained and equipped with materials oriented to his needs and when he and his family continue to receive services such as those provided in the Head Start program, he will continue to accelerate developmentally.

This interpretation has been favored by researchers investigating Head Start programs (e.g., Cawley, 1968) and by researchers reporting follow-up studies of other preschool programs. Cawley writes:

> . . . The tragedy rests in the fact that the overall developmental pattern of these youngsters is so replete with deficits. Society's present course is predicated upon the notion that Head Start will enable these youngsters to catch up. If they don't, then failure in the traditional public school curriculum, often based upon chronological age expectancies for performance, seems obvious. . . . We need to construct a comprehensive system of learning for

> these children. This would entail a number of research and demonstration efforts that would produce successful intervention programs. These would be gradually amalgamated and extended upwards. (pp. 60-61)

Karnes (1969) summarizing a three-year study of children in traditional, ameliorative, and direct verbal training preschool groups comments:

> In spite of the disappointments of some of the longitudinal data, a major accomplishment of this study remains: serious learning deficits of the disadvantaged children in the Ameliorative and Direct Verbal groups were eliminated during the preschool year. In the Direct Verbal group, where extensive special programming was sustained over a two-year period, continued growth occurred. . . . The deterioration in language and intellectual functioning which occurred at the termination of intensive programming demonstrated the need for continued intervention characterized by low pupil-teacher ratios which makes possible the interaction necessary for language development and provides the opportunity to design and implement tasks which will achieve specific goals. (pp. 25-26)

Blaming the school system for the failure of either the increment or acceleration to be sustained seems a plausible and popular interpretation.[4] While a cumulative decrease in academic achievement for low-income children within the school years has been well documented, however, a cumulative increase related to an integrated and continuous preschool and school intervention program has not. There has, in fact, been no experimental test of the five alternate hypotheses that would provide a firm basis for conclusions regarding the effects of sequencing, of density of Head Start children, or of various "optimal" primary environments. We simply do not know what accounts for the often reported "levelling off" phenomenon nor do we really know what kinds of preschool and primary school programs may offer the greatest durability of achievement.

In summary, the available data appear to indicate that there is an immediate impact of Head Start and other preschool programs but we know little about to what this impact may be ascribed or the circumstances under which both change and final levels of attainment may be maximized. Second, the children who have not attended Head Start and other preschool programs tend to catch-up in primary school with those who do attend but we know little about to what this "levelling off" effect may be ascribed or the circumstances under which continued development may be maximized. Head Start can not undertake the exploration of all the alternative explanations for both immediate and "levelling off" phenomena. We have, however, begun to test the conditions under which the cumulative impact of preschool and primary school interventions may be greatest.

An adequate test of the cumulative impact hypothesis is likely to require (1) reduction in the diversity of Head Start programs, (2) investigation of several well-implemented but contrasting approaches, (3) provision of coordinated preschool and primary school experiences that continue each educational approach and other Head Start components (e.g., nutrition, health, parental involvement) through at least the third grade, (4) adopting a long-range evaluation strategy that includes several cohorts followed from Head Start through the primary grades, with additional post-program follow-up after the third grade watershed, (5) designs that provide suitable comparisons for sequencing the interventions across approaches, and (6) measurement of process variables, of criterion-specific variables for each of the approaches, and of diffusion variables.

A planned variation study to be conducted with the cooperation of Follow-Through began this July as a small-scale, experimental effort which in its first year provides comparison across eight approaches implemented in two communities each, with either in-community or similar community "regular" Head Start comparison classes. The 1969 study offers comparison of two groups: children attending both sponsored Head Start and sponsored Follow-Through classes and children attending "regular" Head Start and "regular" primary schools. We hope to expand planned variation in September 1970 to a design permitting in addition a test of the impact of the programs when regular Head Starts are followed by sponsored programs in public school and when the Head Start sponsored programs are followed by "regular" public school experiences.

Considering the evidence now available, we believe that the assumptions on which Head Start was based are still tenable: that from birth through six years of age are important years in human development; that children of the poor generally have not had the experiences and opportunities that support maximum development during this period; that effective programs for these children must be comprehensive including health, nutrition, social services and education; that for their own and their children's benefit, parents should be deeply involved in the design and implementation of local programs; and that a national child development program can focus attention on the needs of preschool and elementary school children from low-income families, and, through continued review of program effectiveness, stimulate local institutions to do a better job of meeting these needs.

The issue regarding the validity of these assumptions is one of inference: Head Start evaluations have tried to locate sources of variation in programs which may affect child development in addition to those of an administrative nature of which we are well aware without evaluation studies. The average impact accruing from all sources of variation—implementation and approach—is another, and different question. From my point of view extensive program revision based on rejection of the assumptions should be deferred until implementation and approach can be distinguished in evaluations.

The problems I have described in Head Start implementation and in evaluation were not introduced to argue against the efforts to measure program impact or to urge moratoriums on program experimentation or an efforts to upgrade program quality. What they may illustrate is the need for careful instrumentation and most particularly for research designs that will explicate interrelationships among program and child variation; the need to study long-term interventions; and the need to avoid quick judgments about Head Start and compensatory education—either favorable or unfavorable—on the basis of data from a program and an art still in their early childhood.

FOOTNOTES

1. Prior to FY '69 many programs were not in operation for the full period due to funding uncertainties and time required for grant application development and grant processing. In 1965-66, most of the relatively few pilot programs were in operation about four months by June 1966. Most of the 1966-67 programs were in operation for less than six months by June 1967. About half of the 1967-68 programs would have been in operation less than six months by June 1968, and of the others funded by continuation grants, perhaps as many as one-fourth were cut short or actually closed down in mid-year for varying periods due to funding cut-backs. Almost all of the 1968-69 programs have operated on continuation funds, which means this period is the first in which the national evaluator could be reasonably certain that the program selected for study would be in operation for the full funding period. The summer programs after 1965 have had relatively uneventful funding histories, although later-than-desired receipt of grants may have affected recruitment and training to an unknown degree. Considering "implementation" as defined in the relatively simple funding pattern, the funding histories of full year and summer programs represents a hidden hazard to evaluation efforts.

2. There is another hidden hazard for the evaluator who seeks to design a pre/post study comparing Head Start and non-Head Start children. Although the program was originally intended to provide a preschool experience for children entering the regular school system in the following year, the actual age of enrollment varies from 3 to 6 1/2 years and many children attend Head Start for two or more years. In addition, due to the conversion of Summer to Full-Year programs, about 30% of a random sample of Head Start children from 1966 on may be expected to have had previous Head Start experience. Another factor is that about 50% of the children now have older siblings who have attended Head Start and/or other preschool programs. The number of new subjects available for an evaluation study is likely to be small, and the analyses would require careful documentation of the child's and the family's history. Location of comparison groups is a third problem. In many rural areas, almost all eligible children attend Head Start leaving only families who are unreachable or ineligible for comparison. In many urban areas only some of the eligible children are served. Horizontal diffusion, older sibling attendance, and participation in other social action programs may mean, however, that although the target "control" child has not participated in Head Start, the family has either been directly affected by Head Start or lives in an area of high concentration of social action programs. Documentation of previous family and child experiences is of particular importance for the urban comparison child; documentation of socio-educational status is of particular importance for rural "controls."

3. It seems clear that scores do not change under all conditions of re-testing and that even multiple testing does not always evoke changes. Datta, O'Keefe and Blanton (1969) compared the average class scores (pre/post) for the same subjects in three treatment groups: test/retest after nine months, monthly testing without feedback to the teachers, and monthly testing with feedback to the teachers. Eight Full-Year Head start classes were assigned at random in September to each of the three groups. The median 18 point gain for the test/retest classes (May PPVT, 85.5 vs September PPVT, 67.9) was not significantly different from that of the two groups of classes tested repeatedly. Gains ranged from class medians of -5 points to class medians of 31 points, with substantial within-class homogeneity in initial and final scores.

4. Christopher Jencks' (1969) reanalyses of the Equality of Educational Opportunity study would seem to offer little encouragement for this hypothesis. Jencks concluded: "My analysis has been confined to what I have described as 'natural experiments,' i.e., variations between schools in the urban North in 1965. An analysis of this kind can tell us little about the consequences of what we might call 'unnatural experiments,' i.e., policies and programs which were not being tried in northern urban schools at that time. Those who argue for the benign effects of such radical innovation—and I am among them—should be troubled by the political difficulty of achieving such innovation on a massive scale. But we need not be troubled by the EEO survey evidence. That survey merely showed that the kinds of innovations which progressive school administrators and lay-boards of education have struggled to achieve in the past (e.g., more money, smaller classes, better trained teachers) would make little difference." (pp. 50-51)

The validity of this conclusion depends in part on whether natural variation yielded any (or a sufficient number) of instances of the innovations that educators have struggled to achieve. One might begin by considering the means and standard deviations of measures of the three innovations which Jencks cites: per pupil expenditures, children per teacher, and two indicators of better trained teachers.

The mean number of pupils per teacher was 28; the standard deviation was 4. (p. 68) If the distribution were symmetrical, the range would be from 16 to 40 children per teacher. At best, a 1:16 ratio of children per adult is not what innovative educators have meant. Although it might be considered better than 1:28 or 1:40, it is not the 1:5 ratio of many innovative preschool programs or the 1:1 such educators as Palmer advocate.

The average per pupil expenditure was $253 a year with a standard deviation of $49. (p. 67) The six standard deviation range would be from $106 to $400 per pupil. The average per pupil expenditure for Full-Year Head Start is about $1,000; the average per pupil expenditure

for northern private schools is about $1,300. The range of per pupil expenditures for the EEO sample of northern schools may include disaster areas at an average of $106 per year, but it is probably not what progressive educators mean by more money at the upper end of about $400 per year.

The average teacher placed the quality of her college at the 27th percentile; the standard deviation was 7 percentile points. (p. 71) The six standard deviation range of teacher estimate of college quality would be from the 6th percentile to the 48th percentile. Again, this doesn't seem to be what progressive educators would mean by high-quality preparation.

It seems likely that until the EEO data are analyzed in something like an analysis of variance model with groups based on more absolute definitions of quality, conclusions about the impact of school quality and innovations on child achievement may be misleading.

REFERENCES

Alexander, Theron. "Changing the Mental Ability of Children in the City." Philadelphia: Temple University, 1968.

Allerhand, Melvin E. "Final Report: Impact of Summer 1965 Head Start on Children's Concept Attainment During Kindergarten," 1967. (ED 015-733)*

Bates, Barbara. "Project Head Start, 1965-67. A Descriptive Report of Programs and Participants." Office of Child Development, U.S. Department of Health, Education, and Welfare, September 1969.

Beller, Kuno E. "Project II: A Study of Cognitive and Social Functioning," 1967. (ED 025-310)*

Beller, Kuno. "Study I: Use of Multiple Criteria to Evaluate Effects of Early Educational Intervention on Subsequent School Performance," 1968.

Berger, Stanley I. "Development of Appropriate Evaluation Techniques for Screening Children in a Head Start Program - A Pilot Project," 1965. (ED 015-006)*

Berlin, Irvin N. "Report of King County, Washington Regional Research Project." Seattle: University of Washington, December 1965.

Bittner, Marguerite L. and Rockwell, Robert E. "An Evaluation of the Pre-School Readiness Centers in East St. Louis, Illinois," 1968. (PS 001-277)*

Cicirelli, Victor G., Cooper, William H., and Granger, Robert L. "The Impact of Head Start: An Evaluation of the Effects of Head Start and Children's Cognitive and Affective Development." Westinghouse Learning Corporation, June 12, 1969. OEO Contract No. B89-4536.

Cawley, John F. "An Assessment of Intelligence, Psycholinguistic Abilities and Learning Aptitudes Among Pre-School Children," 1966. (ED 014-323)*

Cawley, John F., Burrow, Will H., and Goodstein, Henry A. "An Appraisal of Head Start Participants and Non-Participants: Expanded Considerations on Learning Disabilities Among Disadvantaged Children," 1968.

---

*Documents with an ED or PS number are abstracted in "Research in Education," a monthly publication of the Office of Education. See this publication for information on how to order copies through the ERIC Document Reproduction Service.

Cline, Marvin G. "An Evaluation and Follow-up Study of Summer 1966 Head Start Children in Washington, D. C.," 1968. (ED 020-794)*

Chesteen, Hilliard E. "Effectiveness of the Head Start Program in Enhancing School Readiness of Culturally Deprived Children," 1966. (ED 020-771)*

Chorost, Sherwood B., Goldstein, Kenneth M., and Silberstein, Richard. "An Evaluation of the Effects of a Summer Head Start Program," 1967. (ED 014-327)*

Coleman, James et al., Equality of Educational Opportunity. Washington: U.S. Government Printing Office, 1966.

Datta, Lois-ellin, O'Keefe, Ruth Ann, and Blanton, Bradley. "Changes Associated with Three Conditions of Observation in a Sample of Children Attending a Full Year Head Start Program." Mimeo, 1969.

Di Lorenzo, L. T., Salter, Ruth, and Brady, J. "Pre-kindergarten Programs for the Disadvantaged: A Third Year Report on an Evaluation Study," December 1968.

Eisenberg, Leon et al. "Final Report: A Series of Reports and Papers," 1966. (ED 020-773)*

Faust, Margaret. "Five Pilot Studies Concerned with Social-Emotional Variables Affecting Children in Head Start," October 1969.

Gray, Susan and Klaus, Rupert. "The Early Training Project: A Seventh Year Report." Nashville: George Peabody College for Teachers, 1969.

Grotberg, Edith. Review of Research: 1965 to 1969. OEO Pamphlet 1608-13, 1969.

Harding, John. "A Comparative Study of Various Project Head Start Programs," 1966. (ED 019-987)*

Hess, Robert D. et al. "Techniques for Assessing Cognitive and Social Abilities of Children and Parents in Project Head Start," July 1966. (ED 015-772)*

Hillery, Milton C., Lingren, Ronald H. and Remstad, Robert C. "A Descriptive Study of Cognitive and Affective Trends Differentiating Selected Groups of Pre-School Children," 1969.

Hodes, Marion. "An Assessment and Comparison of Selected Characteristics Among Culturally Disadvantaged Kingergarten Children Who Attended Project Head Start (Summer Program 1965), Culturally Disadvantaged Kindergarten Children Who Did Not Attend Project Head Start, and Kindergarten Children Who Were Not Culturally Disadvantaged," 1966. (ED 014-330)*

Hodges, Walter L., McCandless, Boyd R., and Spicker, Howard H. "The Development and Evaluation of a Diagnostically Based Curriculum for Preschool Psychosocially Deprived Children," 1967.

Holmes, Douglas and Holmes, Monica. "An Evaluation of Differences Among Different Classes of Head Start Participants," 1966. (ED 015-012)*

Horowitz, Frances and Rosenfeld, Howard. "Comparative Studies of a Group of Head Start and a Group of Non-Head Start Preschool Children," 1966. (ED 015-013)*

Hubbard, James. "An Exploratory Study of Oral Language Development Among Culturally Different Children," 1967. (PS 000-828)*

Hunt, J. McV. "Comments on Jensen," Harvard Educational Review, Spring 1969.

Hyman, F. A., and Kliman, D. S. "First Grade Readiness of Children Who Have Had Summer Head Start Program," The Training School Bulletin. Vineland: American Institute for Mental Studies, 63, 1967, 163-167.

Jencks, Christopher. "Some Natural Experiments in Compensatory Education." Mimeo: paper presented at SRCD meeting, April 15, 1969.

Jensen, Arthur R. "How Much Can We Boost IQ and Scholastic Achievement?" Harvard Educational Review, 1968, 39, 1-123.

Jensen, Judith and Kohlberg, Lawrence. "Report of A Research Demonstration Project for Culturally Disadvantaged Children in the Ancona Montessori School," 1966. (ED 103-014)*

Johnson, Henry S. "The Effects of Montessori Educational Techniques on Culturally Disadvantaged Head Start Children," 1965. (ED 105-009)*

Kagan, Jerome. "Comments on Jensen," Harvard Educational Review, Spring 1969.

Karnes, Merle B., Teska, James A., and Hodgins, Audrey S. "A Longitudinal Study of Disadvantaged Children who Participated in Three Different Preschool Programs," 1969.

Krider, Mary A. and Petsche, Mary. "An Evaluation of Head Start Pre-School Enrichment Programs As They Affect the Intellectual Ability, the Social Adjustment, and the Achievement Level of Five-Year-Old Children Enrolled in Lincoln, Nebraska," 1967. (ED 015-011)*

Lamb, Howard E., Ziller, Robert C., and Maloney, Alan W. "The Development of Self-Other Relationships During Project Head Start." Newark: University of Delaware, 1965.

McDill, Edward L., McDill, Mary S., and Sprehe, J. Timothy. "An Analysis of Evaluations of Selected Compensatory Education Programs." Paper presented at Evaluation of Social Action Programs Conference, May 2-3, 1969. Mimeo.

Miller, James O. Diffusion of Intervention Effects in Disadvantaged Families. ERIC Occasional Paper, 1968.

Miller, James O. Review of Selected Intervention Research with Young Children. ERIC Occasional Paper, 1968.

Morris, Bernice and Morris, George. "Evaluations of Changes Occurring in Children Who Participated in Project Head Start," 1966. (ED 017-316)*

Muse, Vernon C. An Assessment of the Effects of Head Start Training on Intelligence and Achievement of a Selected Group of First Grade Students. Doctoral Dissertation, Mississippi State University. June 1968.

Nalbandian, Myron K. "Evaluation of the Providence Head Start," November 1968.

Nimnicht, Glen, Meier, John, and McAfee, Oralie. "Interim Report: Research on the New Nursery School. Part I. A Summary of the Evaluation of the Experimental Program for Deprived Children at the New Nursery School Using Some Experimental Measures," December 1967.

Ozer, Mark N. "The Effects of Neurological and Environmental Factors on the Language Development of Head Start Children: An Evaluation of the Head Start Program," 1965. (ED 017-317)*

Porter, Philip J. et al. "Evaluation of Head Start Educational Program in Cambridge, Mass." 1965. (ED 013-668)*

Pierce-Jones, John et al. "Outcomes of Individual and Programmatic Variations Among Project Head Start Centers, Summer 1965," 1966. (ED 014-325)*

Sigel, Irving E. and McBane, Bonnie. "The Relationship Between Cognitive Competence and Level of Symbolization Among Five-Year-Old Children." In J. Hellmuth (ed.), Disadvantaged Child, Vol. 1, 1967.

Soule, Allen. "Northfield, Vermont: A Community Depth Study," 1965. (ED 018-245)*

Sprigle, H., Van De Reit, V., Van De Reit, H., and Sprigle, J. "The Learning to Learn Program: A Sequential Approach to Early Childhood Education and a Study of Its Effectiveness." Report to Carnegie Foundation, 1969.

Steglich, W. G., Cartwright, W. J., and Allen, Shirley. "Report of the Effectiveness of Project Head Start, Lubbock, Texas (Summer 1965)," 1967. (ED 019-131)*

Temp, George and Anderson, Scarvia B. Project Head Start Summer 1966. Final Report Section Three: Pupils and Programs, 1967. (ED 018-248)*

Waller, David and Connors, Keith C. "A Follow-up Study of Intelligence Changes in Children Who Participated in Project Head Start," 1968. (ED 020-786)*

Wolff, Max and Stein, Annie. "Study I: Six Months Later. A Comparison of Children Who Had Head Start Summer 1965, with Their Classmates in Kindergarten (A Case Study of Kindergartens in Four Public Elementary Schools, New York City)," 1966.

Zigler, Edward and Butterfield, Earl C. "Motivational Aspects of Changes in IQ Test Performance of Culturally Deprived Nursery School Children," Child Development, 1968, 39, 1-14.