DOCUMENT RESUME

ED 104 518 PS 007 704

AUTHOR Smith, Marshall S.
TITLE Findings of the Second Year of the Head Start Planned Variations Study.
SPONS AGENCY Office of Child Development (DHEW), Washington, D.C.
PUB DATE Sep 74
NOTE 53p.; Paper presented at the Annual Meeting of the American Psychological Association (82nd, New Orleans, Louisiana, August 30-September 3, 1974)

EDRS PRICE MF-$0.76 HC-$3.32 PLUS POSTAGE
DESCRIPTORS Cognitive Development; *Compensatory Education Programs; *Curriculum Evaluation; *Demonstration Programs; Disadvantaged Youth; Educational Assessment; Goal Orientation; Intelligence Quotient; Longitudinal Studies; *Measurement Instruments; *Preschool Children; Program Evaluation; Psychometrics; Teaching Models
IDENTIFIERS *Head Start Planned Variations; HSPV

ABSTRACT

This paper focuses on a review of data collected from the second wave of a three wave longitudinal study (Head Start Planned Variations-HSPV) designed to assess the relative impacts of a variety of preschool curricula. Head Start children formerly in the HSPV program were tested when they entered selected Follow Through schools and throughout their early school years until they completed Follow Through at the end of the third grade. The study design is reviewed and tests and measures used during the 1970-71 evaluation period are briefly described. Problems in designing assessment instruments for this age group in both cognitive and "non-cognitive" areas are identified. A consideration of the general methodological implications of these design problems is followed by a summary of conclusions on the three main effect questions: (1) overall effects of the Head Start experience; (2) differences between Planned Variation and conventional Head Start programs; and (3) differential effects of the eleven Planned Variation Models (brief descriptions of these are included). Also considered is the question of whether the effects of the various preschool curricula are sensitive to differences in child characteristics. Conclusions address the problem of making curricular evaluations attend more to goals than to psychometric characteristics. (ED)

ED104518

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRO
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIGIN
ATING IT. POINTS OF VIEW OR OPINIONS
STATED DO NOT NECESSARILY REPRE
SENT OFFICIAL NATIONAL INSTITUTE OF
EDUCATION POSITION OR POLICY

FINDINGS OF THE SECOND YEAR OF THE HEAD START PLANNED VARIATIONS STUDY *

Marshall S. Smith
Harvard University (on-leave)
National Institute of Education
September 1974

* Prepared for the 1974 National Convention of the American Psychological Association. Support for the research reported in this paper was provided by Grant #1926, Office of Child Development, U.S. Department of Health, Education and Welfare. The conclusions and recommendations in this paper are those of the grantee and do not necessarily reflect the views of any federal agency.

00002

I. Background and Introduction

During the early months of 1969 the U. S. Office of Child Development planned a three wave longitudinal study designed to assess the relative impacts of a variety of preschool curricula. The study was called Head Start Planned Variation (HSPV) and began in the fall of 1969. Plans called for the systematic assignment of a number of pre-specified curricula, each to two or more sites throughout the country. Selected sites were to meet three criteria.

First, each site was to contain an on-going Head Start Program. No funds were allocated for serving children other than those already being served by Head Start.

Second, each site was to draw participant children from a preschool population living largely within the attandance area of a school or schools where older children attended a Follow-Through program.* By fall 1969, most Follow-Through schools had adopted one of a number of well-defined educational curricula. These programs were being evaluated by the Office of Education. Children entering selected Follow-Through schools during the years 1969-1972 were to be tested at entrance and longitudinally followed and tested until they completed Follow-Through at the end of the third grade.

---

* Follow-Through is an intensive early elementary (K-3) compensatory program designed to enrich the experiences of economically poor children -- particularly poor children who have had Head Start experiences. Originally intended to be a national program, Follow-Through was designated as an experimental effort in 1968, one year after it was initiated. By 1969 there were over 170 school districts with Follow-Through programs.

BEST COPY AVAILABLE

Third, each selected Head Start site had to adopt the curriculum model being used in the Follow-Through schools in its area. Aid in implementing the models was provided by consultants responsible to the original architects of the models. In addition, extra funds for purchasing equipment and for hiring teacher aides were provided to the selected Head Start classes. Overall, the cost of implementing the Planned Variations model was estimated to be $350.00 per child above the cost of conventional Head Start (see McMeekin, 1973). Since many of the Follow-Through curricula were adopted from programs originally designed for pre-schools, the use of them in Head Start programs was appropriate.

The design of the Planned Variations study called for children in all three waves to be tested at the beginning and end of their Head Start experience. Following Head Start, the children would enter the Follow-Through program in their community and be evaluated at the beginning and throughout their Follow-Through experience. Records of the Head Start and Follow-Through evaluation could then be linked. The linkage would provide data for a longitudinal assessment of the combined pre-school and early elementary experiences of the Planned Variation children.

Testing was also planned for other Head Start children in every Planned Variation site. These children would attend Head Start classes without a designated curriculum component and would serve as a local comparison group in the study of the Planned Variations Head Start classes. With some exceptions this strategy was followed for all waves of the Planned Variations study. The comparison children were also to be included in the Follow-Through evaluation.

Three sets of progress reports on Planned Variations were planned: first at the end of the Head Start experience for each of the three waves of children; second, at various times during the Follow-Through experiences of the three waves; and third, in 1976, after the third wave of Head Start children had completed Follow-Through. A preliminary report on short-term effects in the first wave (1969-1970) was prepared in 1971 by the Stanford Research Institute (SRI) for the Office of Child Development.

This paper focuses on data from the second wave of Head Start Planned Variations. SRI collected the data for the second wave and four reports on this data were prepared by the Huron Institute. One report considers the process and success of implementing the Planned Variation curricula in the various sites.

A second report presents a detailed description and analysis of the different measuring instruments used in all three waves of the Planned Variations study. (Walker et. al., 1973). We will refer occasionally to this report and will briefly summarize some of its major conclusions. The third and fourth reports in this series each consider the short-term effects of the various curricula on cognitive outcomes of children. One report (Smith, 1973) focuses on three questions:

1. What are the short term effects of an Head Start experience on children?
2. Are there discernable differences between the effect on children of a Head Start Planned Variation experience and a conventional Head Start experience?
3. Do Planned Variation models differ in their effects on Head Start children?

In all instances the measured effects are narrowly defined. Specifically, we are concerned with three measures of cognitive achievement, one measure of intelligence and one measure of motor control. Although other measures were taken, none were considered to have adequate psycometric properties. Conclusions of this report are presented here in considerable detail.

The fourth report (Featherstone, 1973) systematically explores the possibility that different characteristics of children interact with particular curricula to produce different outcomes. This report uses the data from the first cohort to generate hypotheses and tests these hypotheses on data from the second cohort. It focuses on only two tests -- the Stanford-Binet and the Pre-School Inventory (PSI). A number of the major conclusions of this report are summarized here.

This paper is divided into four additional sections. In section II we describe the design of the study and briefly describe the tests and measures. The main conclusions of the report by Walker et al. are considered in this section. Section III considers the three main effect questions analyzed by Smith and Section IV focuses on a number of the major findings of the study of interactions. Section V contains some brief conclusions.

II. Study Design and Measures

A. Study Design

Thirty-seven sites used Planned Variations curricula in 1970-71. Twelve curricula models were represented.* There were comparison classes at 14 of the 37 sites and at seven (off-site) locations not having Planned Variation classes. Table I displays this information. Columns 1, 2 and 3 of Table I show the names of the twelve curricula models, a site code for each site, and the location of the site. The model and site codes were assigned in the Follow-Through evaluation (with the exception of the Enabler model) and contain no information other than identification of model and site. The Enabler sites are unique to the Head Start Planned Variations study.

Column 4 indicates the testing level which is described later. Column 5 of Table II-1 shows the year of entry of the site into the Planned Variations study. Fifteen of the 37 sites were also in Planned Variations in 1969.

---

* For description of the models see Appendix I. As indicated in the appendix the models vary on a number of different dimensions.

TABLE I

HEAD START COMMUNITIES 1970-71

1	2	3	4	5	6	7	8
RRICULUM MODEL SPONSOR	SITE CODE	SITE COMMUNITY	TESTING LEVEL	YEAR SITE JOINED STUDY	# OF PV CLASSES #	TESTED	NUMBER OF COMPARISON CLASS
	02.02	Buffalo	I	70	11	4	
Nimnicht	02.04	Duluth	III	70	9	4	
(Far West	"	St. Cloud	III	70			2
Laboratories)	02.05	Fresno	III	70	4	4	
	02.09	Salt Lake	I	69	6		
	02.13	Tacoma	II	70	7	4	
	03.08	LaFayette	III	69	17	4	
Henderson	"	Albany	III	69			4
(Tucson)	03.09	Lakewood	I	69	8		
	03.16	Lincoln	III	70	7	4	
	05.01	Boulder	III	70	4	4	1
	05.10	Tuskegee	I	69	12		
Bank	05.11	Wilmington	II	69	9	4	
Street	"	DeLaWar	II	69			4
	05.12	Elmira	III	70	7	3	3
Becker &	07.03	E. St. Louis	III	69	9	4	4
Englemann	07.11	Tupelo	III	69	4	4	4
(Oregon)	07.14	E. Las Vegas, NM	II	70	5	4	
	"	W. Las Vegas, NM	II	70			4
	08.02	Oraibi	III	69	7	--	
Bushell	"	Acoma	III	69			4
(Kansas)	08.04	Portageville	III	69	4	4	4
	08.08	Mounds, Ill.	II	70	5	4	2
	09.02	Ft. Walton Bch.	III	69	5	4	
	"	Pensacola	III	69			3
Weikart	09.04	Central Oz	I	69	16		
(Hi-Scope)	09.06	Greeley	III	70	4	4	3
	09.10	Seattle	II	70	6	4	3
	10.01	Jacksonville	I	69	3		
Gordon	10.02	Jonesboro	III	69	3	3	3
	10.07	Chattanooga	III	70	9	4	4
	10.10	Houston	II	70	7	4	4
	11.05	Washington	III	69	5	4	4
EDC	11.06	Paterson	II	70	4	3	4
	11.08	Johnston Co.	III	69	6	4	4
Pittsburgh	12.03	Lock Haven	III	70	7	4	
(IPI)	"	Mifflenburg	III	70			4
REC	20.01	Kansas City	III	70	8	4	
N.Y.U.	27.01	St. Thomas, VI	I	70	4	-	4
	27.04	Billings	II	70	5	4	
	27.05	Colorado Spr.	II	70	6	4	
Enablers	27.03	Bellows Falls	II	70	6	4	
	27.02	Newburgh	I	70	8		
	27.01	Puer Rico	I	70	6		

BEST COPY AVAILABLE

Column 6 shows the number of Planned Variation classes in the site and Column 7 shows the number of tested classes. Column 8 shows the number of comparison classes all of which were tested. Blanks in Column 6 indicate that the site was an "off-site comparison" site. Note that the off-site comparisons are paired with Planned Variation sites and are given the same site code as a Planned Variation site. Blanks in Column 8 indicate that there were no comparison classrooms at that site.

There are a number of things to note about the design. First, nine of the 12 models have at least 2 sites and 5 of the 9 have 3 sites. Although it would have been preferable to have greater replication, the very fact that some systematic replication did occur strengthens inferences about model effects.

Second, the study is not confined to one region of the country. Although there is some regionality associated with some of the various models, e.g., the Gordon model sites are all in the South; by and large, each model is being tried at a variety of locations.

Third, a number of sites do not have comparison classes in their same location. This tends to complicate analysis of the data.

B. Measures Used in the Evaluation

The level of testing indicated in Column 4 of Table I refers to the evaluation activities carried out in the sites. Due primarily to economic constraints, not all sites were tested on all measures. There were three levels of evaluation activities -- I, II, III. Table II describes the

activities included at each of the three levels. Level I is the most primitive. Eight Planned Variation sites fall in this category. No comparison sites are in Level I. No data gathering at this level involved the children. Teachers completed demographic information forms and filled out the California Social Competency Scale for each child in their classrooms in both the Fall and Spring. Also, teachers and teacher aides responded to a questionnaire requesting information about their own backgrounds, teaching experiences and attitudes. In addition, Sponsor and Head Start Directors rated the level of implementation in the classrooms in every site.*

All data collected at Level I was also collected at Level II. In addition, three other sets of data were gathered at Level II. Classroom observations were made in both the Fall and Spring by observers using the SRI Classroom Observation Instrument. All children in tested classrooms were administered the Basic Test Battery in both the Fall and Spring. Four tests were included in the Basic Battery -- the Caldwell Preschool Inventory, NYU Booklet 3D, (a 17 item test of achievement -- pre-math, pre-science and linguistic concepts), NYU Booklet 4A, (an 18 item test of achievement -- letters, numerals, prepositions) and the Motor Inhibition Test. Children of Black and Spanish origin, whose parents agreed, were administered a test assessing their knowledge of their ethnic heritage. Eleven Planned Variation sites were classified as Level II. Of the eleven, six had comparison classes also tested at Level II.

---

*See appendix for a short description of the measures used in the study.

TABLE II

Three Levels of Planned Variation Evaluation Activities

Level of Evaluation	Data Collection Period		
Level I	Fall		Spring
1) Teacher completed classroom information forms -- for child demographic data	X		X
2) Teacher completed California Social Competency Scale -- one for each child	X		X
3) Sponsor ratings of Level of Implementation	X	X	X
4) Head Start Directors ratings of Level of Implementation	X	X	X
5) Teacher and Teacher Aide survey			X
Level II (includes all activities in Level I and the following)			
6) Classroom observations	X		X
7) Basic Child Test Battery a. Preschool Inventory b. NYU Book 3D c. NYU Book 4A d. Motor Inhibition Test	X		X
8) Child completed Ethnic Heritage Test	X		X
Level III (includes all activities in Levels I and II and the following)			
9) Stanford-Binet testing on random one-half of children in all tested classes	X		X
10) 8-Blocks Sort Task -- given to other random one-half of children in all tested classes			X
11) Parent Interviews -- administered to parents of children taking the 8-Block Sort Task			X
12) Intensive Case Studies (U. of Maryland)			X

Level III sites had all data collection activities carried out in Level I and Level II sites and, in addition, four other activities. Randomly chosen, one-half of the children in each tested Level III classroom were administered the Stanford-Binet in both the Fall and Spring. The same children received the test both times. The children in the other half of the class -- along with one of their parents or guardians -- were administered the 8-Block Sort Task in the Spring. Additionally, the parents or guardians of the children in this group completed a parent questionnaire which asked about attitudes toward Head Start, their child, and the Planned Variation model used in their child's classroom. Eighteen Planned Variation, ten on-site comparison, and five off-site comparison sites were assigned to Level III.

In an extensive review of the measures and data collection activities of the three cohorts of HSPV Walker et al. (1973) reach four general conclusions. These conclusions are as appropriate to the 1970-71 data as they are to the entire study. In brief they are:

1). A large number of measures exist which purport to assess the cognitive characteristics of preschool youngsters. Many of these measures are apparently well constructed and show high internal and test-retest reliability. By and large, however, little data exist concerning the long-term predictive validity of the measures. In general, data on predictive validity are limited to correlations between the tests given at preschool time and similiar tests given in the early elementary years. These correlations are generally moderate falling in the 0.40 - 0.50 range at the outside. One exception to this is the Stanford-Binet which shows

a fairly stable relationship to itself from the preschool years to late adolescence -- the correlation between the Binet at age six and the Binet at age 18 is in the range of 0.60 -0.70 (see Jencks et al., 1972). Moreover the Binet seems to be fairly strongly related to eventual years of schooling completed (we estimate that r=0.42 between IQ at age 5 and eventual years of schooling). The relationship to such future outcomes as occupational status and income is considerably less. An important drawback to these data, however, is that they represent relationships estimated in the absence of effects of preschools. It is entirely possible, for example, that preschool treatments which affect the IQ level of young children also change the predictive relationship between early and late IQ. Even in the absence of data answering these questions, however, the existing cognitive measures are psychometrically far stronger than the measures of "non-cognitive" areas.

2). Measures in the "non-cognitive" areas should be uniformly classified as experimental. With only a few exceptions little data exist as to their reliability or validity. Moreover, they are characterized by a lack of a clear relation to theory and by severe problems in administration. (For an extensive review of "non-cognitive" measures for preschool age children see Walker, in press). In every instance we classify the non-cognitive measures used in HSPV as inappropriate for use as summative evaluation instruments. This is particularily unfortunate in this study given the diversity of goals of the various sponsors in HSPV and represents a severe drawback in the evaluations.

3). By and large comprehensive norms and assessments of the psychometric characteristics of both cognitive and non-cognitive measures do not exist for

the population of children served by Head Start. We attempted to alleviate this problem for the measures used in HSPV by extensively describing the data and we recommend that future researchers follow a similiar strategy in future evaluations.

4). Over all we found the data collection procedures followed by SRI to be highly competent and productive of generally very high quality data. However, although the data collection procedures were exemplary, the design of the study raised some sticky problems for analyses. These are considered in the next section where findings of the main effects study (Smith, 1973) are summarized.

III. The Main Effects Study

This section contains two parts. First we consider some of the general methodological implications of the design problems mentioned in Section II. We then summarize conclusions to the three main effect questions.

A. Methodological Implications and Strategy for Analysis

The strategy for analysis was dictated in large part by constraints on the study. First, as we noted earlier, we focused principally on the analysis of cognitive growth. To do otherwise would be to seriously overplay the existing data. In doing this we recognize that we are not even attempting to capture the richness of a preschool experience or the largest part of the differences among preschools.

Second, also as described earlier, Planned Variation sites were not randomly assigned to models. Rather, the sites were selected on two criteria unrelated to the requirements for an adequate experimental design, and then given the opportunity to accept or reject the assigned curriculum. Moreover, the local community had control over the specification of which classes within a site were to employ the Planned Variation (PV) curriculum. Since the selection of comparison classes within the PV sites occurred after selection

of the Planned Variation classes, the treatment (PV) and comparison (NPV) classes cannot be assumed to be random samples drawn from the same population. Thus, randomization did not occur at either of the two critical design points -- at the level of assignment of curricula to sites or at the level of assignment of treatment and comparison groups within sites.

The lack of a true experimental design puts the analysis of the data into a never-never land. Had we random assignment of curricula (treatments) to sites, then a comparison of treatments would yield us unbiased estimates. If we had random assignment of classes to PV and NPV groups within sites, then a comparison of the two sets of classes would yield unbiased estimates. If we had two random samples of children from the same population -- one going to Head Start and one not, then estimates of the general effects of a Head Start experience would be unbiased. But we have no random assignment, so all estimates are biased in some unknown fashion. Estimation of effects thus becomes an art instead of a science. There are numerous statistical techniques to help reduce bias (matching, covariance, blocking, crossed designs and standardization techiques). Each may be helpful depending on the adequacy of the structural model we are trying to fit. That is where the essential problem lies, for we have no a priori way of determining which is the best analytic model. Given this state of affairs, we follow Tukey's advice: "As in the famous discussion between Student and Fisher and the interjections by Sir Harold Jeffreys, it may not be a bad thing to use all the allowed principles of witchcraft and not just one set." (Tukey, 1973, p. 112).

We did not use all the principles of methodological witchcraft, but we did use a number. In particular, our strategies for removing bias in the data depend on (1) our choice of a statistical model; (2) our choice of variables; and (3) our assessment of the accuracy with which the data are measured. Different decisions in these areas of judgment lead to the use of different statistical approaches which in turn leads to a variety of estimates of "effects". In a sense, different estimates from different analyses gives us a sense of confidence limits for the reported effects. Such an approach will generally inspire caution in interpretation, for most of the effects found in this study are small and somewhat sensitive to differences in the analytic strategies. On the other hand, large effects which turn out to be robust -- insensitive to variations in analysis methods -- presumably should inspire confidence.

B. Summary of Findings in Main Effects Study

Three main questions were addressed in this study:

-- What are the overall short term effects of a Head Start experience on children?

-- Are there discernable differences between the effects on children of a Head Start Planned Variation experience and a conventional Head Start experience?

-- Do Planned Variation models differ in their effects on Head Start children?

Five measured outcomes were used to assess each question. The PSI is a general standardized achievement test for pre-school children. The NYU Booklet 3D and NYU Booklet 4A are tests of specific achievement areas. The Stanford-Binet is a well known test of general "intelligence". The Motor Inhibition test assesses a child's ability to control motor behavior.

1. Overall Effects of the Head Start Experience

With regard to the question of overall short term effects of Head Start we reached four conclusions:

a. The Head Start experience substantially increased children's test scores on all five outcome measures. On four of the five outcome measures children's scores were estimated to increase "naturally" over the seven or eight months of the Head Start program. Thus, even had the children not been exposed to Head Start, their scores would have risen. For two of these measures (PSI and Booklet 3D) the Head Start experience was estimated to double the "natural" rate of growth. For two other measures (Booklet 4A and the Motor Inhibition tests) the Head Start experience was estimated to better than triple the "natural" rate of growth. Increments attributable to Head Start ranged from 0.26 standard deviations (for the Motor Inhibition test) to 0.82 standard deviations (for the Booklet 4A test). On the fifth measure, the Stanford-Binet, our estimates indicate that the scores of children in this sample

TABLE III

Gains for the total analysis sample on 5 outcome measures. Observed gains are partitioned into two components -- gains attributable to maturation and gains attributable to an Head Start experience. All gains are expressed in individual level pre-test standard deviations.

Test	Observed gain (total)	Attributable to maturation (estimated)	Attributable to Head Start (estimated)
PSI	0.942	0.496	0.446
Book 3D	0.727	0.363	0.364
Book 4A	1.151	0.333	0.818
Motor Inhibition	0.36	0.10	0.26
Stanford-Binet	0.348	-0.296	0.644

TABLE IV

Overall Average Changes in Mean Test Scores for Children in the 1970-71 Final Analysis Sample

2,235 children are represented in the table. Each cell contains the mean gain and the number of children in the group. (Blank cells indicate insufficient N to estimate mean).

Group	Ethnicity	Prior Pre-School Exper.	Entering Grade	PSI Gains	Book 3D Gains	Book 4A Gains	MI Gains	Stanford Binet Gains
				Groups				
1	Mexican-American	No	El 1	$\bar{x}$=12.5 n= 111	2.6 111	4.0 111	0.389 62	-----
2	Mexican-American	No	Ek 2	9.7 106	1.7 106	2.9 101	0.463 21	7.5 40
3	White	No	El 3	10.4 157	2.5 158	5.8 158	0.284 102	7.7 65
4	White	No	Ek 4	12.6 426	2.7 422	3.7 418	0.434 168	5.6 147
5	Black	No	El 5	11.3 252	2.8 250	6.0 250	0.380 115	2.6 129
6	Black	No	Ek 6	12.8 561	2.1 557	2.6 556	0.402 135	6.9 164
7	Mexican-American	Yes	El 7	-----	-----	-----		-----
8	Mexican-American	Yes	Ek 8	-----	-----	-----		-----
9	White	Yes	El 9	7.1 69	1.9 69	5.4 69	-0.001* 46	3.6 31
10	White	Yes	Ek 10	7.9 75	1.8 75	2.7 74	0.460 32	2.2 * 27
11	Black	Yes	El 11	10.2 111	2.5 111	5.1 111	0.210 49	-0.3 * 48
12	Black	Yes	Ek 12	9.0 115	2.2 114	2.6 113	0.347 15	3.4 29
	Total			$\bar{x}$=11.4 n=2,003	2.4 1,993	3.8 1,981	0.36 765	4.7 625
			SD_{gain} =	8.0	2.9	4.2	0.63	11.1
			SD_{pre} =	12.1	3.3	3.3	0.54	13.5
			SD_{post} =	10.3	3.1	4.7	0.57	13.2

*Indicates change is not statistically significant beyond the .05 level

would have "naturally" decreased by about 0.30 standard deviations had they not attended Head Start. The Head Start experience arrested this apparent decrease and further increased Head Start participants' Stanford-Binet scores by roughly 0.35 standard deviations.

b. Children who had a prior preschool experience gained less overall ("natural" + Head Start related growth) than children for whom 1970-71 Head Start was their first year of preschool. This effect held for all outcome measures and for most of the subgroups (see Table IV). Featherstone (1973) reached the same conclusions for the PSI and the Stanford-Binet using a somewhat different analysis strategy. If, however, we allocate the total gains for the two groups of children between "natural growth" and the Head Start experience, we find that the effects attributable to Head Start are roughly equal for children with and without prior preschool experience (see Table V). This indicates that the expected "natural growth" for children with prior preschool experience is less than for children without prior preschool. The prior preschool experience appeared to reduce differences in test scores between children of different ages. In other words, a common preschool experience partially overcomes the effect of age differences among children on the five outcome measures. Some support for this notion comes from the fact that variances on four of the five outcome measures are somewhat smaller at post-test time than at pre-test time. This indicates that differences among children are less at the end

TABLE V

Overall Gain for children in the Planned Variation Study in 1970-71. "Gains" are computed by subtracting an "Expected post score" from an "Observed Post Score." (Blank cells indicate insufficient numbers of children to estimate mean "gains.")

Group	Ethnicity	Prior pre-school exp.	Entering Grade	PSI "Gain"	Book 3D "Gain"	Book 4A	Motor Inhib.	Stanford-Binet
				Gain= 6.7 n = 111	1.8 111	2.6 111	0.14* 63	---
2	M-A	No	Ek	0.40* 106	0.2* 106	1.9 101	0.19 22	9.3 40
3	Wht.	No	El	4.7 157	1.1 158	4.7 158	0.14 103	8.0 65
4	Wht.	No	Ek	4.0 426	1.1 422	1.9 418	0.28 169	7.0 147
5	Blk.	No	El	8.4 252	2.4 250	4.8 250	0.40 116	9.6 129
6	Blk.	No	Ek	6.4 561	1.0 557	1.9 556	0.38 136	10.4 164
7	M-A	Yes	El	---	---	---	---	---
8	M-A	Yes	Ek	---	---	---	---	---
9	Wht.	Yes	El	7.1 69	0.4* 69	3.4 69	0.05* 47	12.2 31
10	Wht.	Yes	Ek	8.8 75	2.4 75	2.7 74	0.47 33	14.5 27
11	Blk.	Yes	El	7.3 111	1.7 111	4.8 111	0.03* 50	8.4 48
12	Blk.	Yes	Ek	-0.7* 115	-0.2* 114	1.1 113	0.47 16	-2.0* 29
			TOTAL	5.4 2003	1.2 1993	2.7 1981	0.26 765	8.7 686
			SD_{gain}	=9.2	2.8	4.5	0.56	12.2

* Indicates gain not statistically significantly greater than zero beyond the .05 level.

of the preschool program than they are at the beginning of the program. Preschools may have a "fan-close" rather than a "fan-spread" effect on children.

c. Children who would enter first grade (E1) directly from Head Start tend to gain more than children who would enter kindergarten (Ek) directly from Head Start on the Booklet 4A, Booklet 3D, PSI and Stanford-Binet tests.* On the motor Inhibition test the Ek children gained more. (The average age of E1 children when they entered Head Start was 65 months-- Ek children were roughly one year younger.) The greater gain for E1 children was most pronounced for the Booklet 4A test and least for the Stanford-Binet (see Table IV). When the gains attributable to Head Start were examined, the effect appears to strengthen, though they are still small for the Stanford-Binet (see Table V). These effects are probably due to a combination of two things. First, the larger gains attributable to Head Start for E1 children on the cognitive measures and particularly the Booklet 4A test (a measure of letters, numerals, and shape names) may be due to older children's advanced academic readiness. Second, there may be a greater interest by Head Start teachers in E1 sites in preparing children for reading and arithmetic.

---

* A number of sites (particularly in the Southeastern areas) do not have kindergartens. In these sites Head Start children generally go directly from Head Start into first grade.

d. There seem to be no consistent differences among Mexican American, black and white children in their Head Start gains on the five outcomes.

A brief mention of the methodological procedures used to arrive at these conclusions is in order. Since we did not have a group of "control" children (children who did not have the benefit of an Head Start experience) our procedures for separating "maturation" from "Head Start" gains relied on natural variations in prescores for children of different ages. The reader is warned to treat these data as rough estimates and to evaluate for himself the assumptions of the procedures. I should note, however, that analysis of the third wave of data (1971-72), when there was a real "control" group indicate that the estimates are generally accurate.*

---

* Since we did not have the appropriate control group in the study -- that is, a group of children who did not attend preschool -- we had to make our estimates of "natural gain" and "gain due to Head Start", by using age variations in the children in the study. Briefly, our procedure was as follows: First, we divided the children into the twelve groups represented in the tables. Second, we took the pre-test scores of these children on all of the five variables we were interested in and ran 60 separate regression equations (1 for each pre-test variable for each of the 12 groups). The pre-test scores were the dependent variables in the regression analyses. The independent variables for each analysis were age, sex, family income, household size, mothers' education and appropriate dummy variables to control for missing data. By using the coefficients for these equations and the original data, we came up with an "expected" pre-score for each child. Within each of the 12 sub-groups, the mean of the expected pre-scores equals the mean of the actual pre-scores.

Our analysis required two more steps. Remember that the procedure controls for the effect of age with the three stratifying variables and their interactions as well as for the variables in the regression equations (sex and family background). Thus, with respect to the relationship between age and the PSI test score, we can argue that the coefficient for age within any given equation reflects the rate of growth for the children within that group -- other reasonable things being equal. Or in other words, other things being equal, we can argue that the main difference between the score of a child entering the program when he is 48 months old and his score if he had entered the program when he was 55 months old, is reflected in the coefficient for age for his particular group. Thus, given this argument, we can estimate an expected post-test score for the child a score that should approximate

2) Differences between Effects of PV and Conventional Head Start Programs*

The second major question regards overall differences in effects for Planned Variation and conventional Head Start programs. Although we considered the question in some detail, we argue that the question has very little importance. For while we might expect there to be differences among PV programs in their effects on the five outcome measures, we have little reason to suspect that there should be systematic differneces between an overall PV effect and an overall effect of conventional Head Start programs. This question, like many total program impact questions, totally obscures systematic differences among treatments.

The sole rationale for studying the question is to determine whether the extra funds allocated to PV Head Start programs had a consistent effect on the measured outcomes. Our conclusion supports the findings of a large number of recent research efforts which have failed to detect any systematic relationship of gross expenditures to variations in outcomes. We conclude there are no differences in effects between the PV programs (taken together) and the Comparison Head Start programs on any of the five outcome measures.

(con't) the score he would have recieved on a post-test. Assume, for example, that a child's estimated pre-score is 40. Also assume that the preschool program is 8 months long and that the coefficient for age is 0.5 -- a gain of 1/2 point/month. Then we can estimate what the hilds' score would be if the programs had no effect. It would be 40 + .5 (8 months) [44]= expected score. Then if post-test score was 48 -- the "gain" due to Head Start would be 4 points.

We followed this procedure: for each child in each of the groups on all five tests, we estimated an expected "post-test" score. We then subtracted this score from his actual post-test score and computed group means. The group means are shown in Table 5. I should note two problems with this procedure. First, there may be an effect on the post-test score created by the fact that the child has taken a pre-test. This is uncontrolled for in this analysis. My guess, however, is that this effect is minimal compared to the effect of the preschool socialization experience on the test-wiseness of the child. Second, the validity of the approach depends upon the quality of the controls -- used both in the stratification or in the regressions. We have to be able to argue that the coefficients for age are unbiased.

*For details on the analysis strategies used to reach these conclusions and the conclusions relating to model-to-model differences see Smith, 1973. As indicated in the text a variety of analytic strategies were used to reach each set of conclusions. These included Analysis of Covariance, matching techniques, the standardization technique used in the assessment of the overall Head Start effects.

3) Differential effects of the Various Planned Variation Models

The third question addresses differences among PV programs in their effects on Head Start children.

Table VI summarizes our findings regarding differential model effectiveness. The eleven PV models are the rows of the table while the five outcome measures are each represented by a column of the table. The cell entries indicate effectiveness relative to the other PV models and to appropriate conventional Head Start classrooms. Five general conclusions may be reached after inspection of this table.

a. We began the study with the expectation that there will be few strong differences among the models in effectiveness as assessed by our five outcome measures. By and large this expectation was realized. Table VI indicates that for each of the outcome measures we have classified the majority of the models as having average effectiveness. Moreover, no model stands out as either more or less effective than the others on more than two of the five outcomes. In the crudest terms there are no overall winners or losers.

b. A second and more tentative expectation was that models which emphasized academic drill combined with systematic reinforcement would be more effective than other models on the four cognitive outcome measures. This expectation was realized only for one of the four cognitive measures. Only for the Booklet 4A measure -- a test assessing knowledge of letters, numerals, and shape names-- is there evidence of greater effectiveness for the models emphasizing

TABLE VI

Summary of Planned Variation Model Effectiveness on Five Outcome Measures

Zero (0) indicates model is of average effectiveness on outcome measure.
Minus (-) indicates model may be of below average effectiveness.
Plus (+) indicates model may be of above average effectiveness.
Double plus (++) indicates model is probably highly effective.

Model	Book 3D	Book 4A	PSI	Stanford Binet	Motor Inhibition
Far West Laboratory	0	0	0	0	0
Arizona	0	0	0	0	0
Bank St.	0	0	0	-	+
Univ. of Oregon	0	+	0	0	0
Univ. of Kansas	0	++	0	0	+
High Scope	+	0	0	++	0
Univ. of Florida	-	0	0	0	0
EDC	0	0	0	0	0
Univ. of Pittsburgh	0	+	0	+	-
REC	-	-	0	+	0
Enablers	0	-	0		+

drill and reinforcement. The University of Kansas model is the clearest example of this finding. We found it to be clearly superior to all of the other models and to the Comparison classes in its effectiveness in raising Booklet 4A test scores. The two other models which emphasize academic drill (University of Oregon and University of Pittsburgh) both appear to be above average in their impact on this test.

On the other cognitive tests there is no indication of special effectiveness of these three models. Only the University of Pittsburgh model on the Stanford-Binet shows an other than average effect. These findings are at some variance with the finding of other researchers in the pre-school area (see Bissell, 1970; and White, S., et al., 1972). These researchers indicated that there may be a general positive effect of structured academic emphasis and drill on cognitive tests. Our data, however, indicate that the effect is specific rather than general. In particular it appears as if this approach may be more effective for imparting information that is easily taught through systematic drill while it is only of average effectiveness in other cognitive areas. Of the four cognitive tests the Booklet 4A test most clearly assesses specific skills. The other tests, particularly the PSI and the Stanford-Binet, assess general information and cognitive functioning.

c. One model clearly stands out as more effective than the others in raising Stanford-Binet test scores. The High Scope PV model appears to increase Stanford-Binet scores by an estimated twelve to fifteen points, roughly 0.9 individual level standard deviations. The average effect of other PV and Comparison models is on the order of three to four points or roughly 0.3 standard deviations. The estimates of these effects are based on analysis of two very different sites. They also replicate the general findings from the 1969-70 cohort (Featherstone, 1973). The effect of the High Scope model is particularly strong in one Southern rural site where the measured average gain is slightly over thirty points. Although we can probably attribute some of the measured gain to tester and regression effects, the "corrected" gain in this site is still on the order of a very substantial twenty points. Preliminary analyses of the item profiles of children in the High Scope sites indicates that the gains apply over the entire range of appropriate Stanford-Binet items (see Butler, 1973).

The particular effectiveness of the High Scope model on the Stanford-Binet does not appear to generalize to the other outcome measures used here. For three of the four remaining tests the model appears to be of only average effectiveness. On the fourth test, Booklet 3D, there is some indication that the High Scope model may be of above average effectiveness but no firm conclusion may be reached from the data.

d. Two of the eleven models (University of Pittsburgh and REC) account for 40% of the 15 cells in Table VI where there is an indication that a model has other than average effectiveness on an outcome measure. Pittsburgh appears above average on the Booklet 4A and Stanford-Binet tests and below average on the Motor Inhibition test. REC appears below average on the Booklet 3D and Booklet 4A tests and above average on the Stanford-Binet. No other model is rated as other than average on more than two of the measures. Three things are common to REC and Pittsburgh. Each uses some form of programmed instruction, each was a first year model in 1970-71, and each has only one site in this study. Although the first two common elements may be important, our inclination is to view the fact that each model has only one site as the principal reason that these models have more than their share of "other than average" effects. This may be due to differential effectiveness of models in different sites or to uncontrolled biases in our data. Whatever the reason, our inclination is to be very skeptical about attributing clear effects to any model with only one site.

e. All models are rated as showing average effectiveness on the PSI test. We had not expected this result since our preliminary analyses of the PSI indicated that it is probably our most reliable measure. In retrospect, however, we suspect that the reason for

the lack of clear differences among models on the PSI is due to the nature of the test itself. The PSI was developed as a general test to assess the overall impact of preschools on children. As such it attemtps to measure a wide range of skills probably rendering it relatively insensitive to particular differences among curricula. Thus it is probably more appropriate to the tasks of assessing the overall average impact of preschools and of individual differences among children.

This brief summary highlights the central findings of the main effects study. Th conclusions regarding the effects of particuliar sponsors held up in a number of different types of analysis. In the next section we consider the question of whether th effects of the various preschool curricula are sensitive to differences in child characteristics.

IV Child Characteristic by Model Interactions

Two general strategies were followed for the analysis of child characteristic by model interactions in the 1970-71 HSPV data. First, Smith, 1973 explored a variety of potential interactions in the context of the main effects study to determine if general conclusions about main effects should be tempered by the understanding that the main effects are only averages and that the programs work somewhat differently for different kinds of people. In keeping with the "conservative" nature of this study the exploration was limited to instances where clear main effects were present. The conclusion from these analyses is clear. While it is certainly true that the models which showed large effects were not equally successful for all examined types of children, major disordinal interactions did not appear. Thus, a model that was effective on-the-average for a certain test was not particularily "effective" for one type of child and particularily "ineffective" for another. This was especially true for the two very

program appeared to produce gains on the Stanford-Binet which exceeded those of almost all of the other programs for all types of children. A similiar conclusion applies to the effectiveness of the University of Kansas with the Book 4A test.

The second type of exploration was carried out by Featherstone, 1973. Her analyses focused directly on the question of child-characteristic by model interactions as a research issue rather than as a mechanism for interpreting main effects. The focus in these analyses was limited to the PSI and the Stanford-Binet. Three general sets of findings emerged from her analyses.

1). Although analyses of interactions in both the 1969-70 and in the 1970-71 suggest that there are a lot of things going on in the data there was little comparability between conclusions that could be reached for the two years. Moreover, it was generally the rule rather than the exception that the two outcome measures reacted differently with regard to the interactions. These findings may be due to a variety of factors including the relative weakness of the 1969-70 data, non-comparability between the samples in the two years, an advanced state of implementation of the models in the second cohort, or a general lack of real consistency in the effects. For an important exception to this generalization, note the third finding below.

2). A focus on the second year data combined with an a priori grouping of the models to add power to the analysis, however, provided a number of very provocative findings. In these analyses two "more directive" or "adult controlled" models (University of Oregon and University of Kansas) were contrasted with three "less directed" and "student initiated" models (EDC, Bank Street and Far West Laboratories A number of findings emerged from these comparisions. The "more directive" models tend to favor: a) children without preschool experience; b) children who initially score on the low end of the distribution (PSI only); c) boys with preschool experien in contrast to girls with preschool experience; and d) younger children. The "less

directed" models favor a) children with preschool experience (PSI only); and b) initially high scoring children. Interactions of SES and ethnicity with model type are generally negligible. While all these findings are not reproduced in both cohorts, they appear to be somewhat coherent and deserve close attention in future studies.

3). One set of findings appeared to replicate in the two cohorts. Associated with the administration of the Stanford-Binet was a behavioral scoring system developed by Hertzig and Birch (Hertzig et al., 1968). Two measures adapted from this scoring system and assessed at pre-test time were used to categorize children. One measure, "passivity" was indicated by a child remaining silent and passive if he did not know the answer to an item. The second measure, "competence" was indicated by a child attempting to answer items he was unfamiliar with and elaborating on correct answers. Children were classified as high or low on these two measures. In both years two general tendencies were discovered: "more competant" and "less passive" children were more successful in the "less directive" models and "less competent" and "more passive" children were more successful in in the "more directive" models. The findings were clearest for the Stanford-Binet thought the PSI data tended to fall in the same pattern. Of crucial importance in interpreting these findings is the fact that the "passive" and "competence" measures were the only directly observed behavioral independent variables used in the interaction study. It is very possible that future studies of child-characteristic by curriculum model interactions will waste their time if they don't focus on behavioral rather than "sociological" variables.

V. Conclusions

A number of the findings of this study are not surprising. Numerous studies have shown that Head Start has a clear and strong short term effect on cognitive outcomes (see e.g., Datta, 1972; or Sterns, 1972). And many more studies of

school effects indicate that it is unlikely that any one curriculum will show a clear advantage over other curricula in the production of cognitive growth (see Jencks, et al., 1972). However, for the effects which do persist in our analyses there is a seemingly clear relation between the goals of a preschool curriculum and the child outcomes "produced" by the curriculum. The success of the academically oriented curricula on the test of letters and numerals, High Scope's success on the Stanford-Binet, and the apparent success of the "directive models" with younger, more "passive" and generally less ready children are the three principle examples.*

These findings, along with the apparent insensitivity of the Preschool Inventory to different curricula, seem to indicate that evaluations of preschool (and school) curricula should attend more to the goals of the curricula than to the psychometric characteristics of the tests. As indicated in the discussion of non-cognitive measures, of course, this is easier said than done. At the least, however, we need to have a diversity of cognitively oriented measures in future assessments. If, for example, we had used only the most "reliable" and widely used test in this study (the PSI), we would have had very little of interest to report.

---

*Analyses of the first and third wave of the Planned Variation data indicate that these two sets of results are not specific to the year 1970-71.

Averch, H.A., et al. How effective is schooling? A critical review and synthesis of research findings. Santa Monica, California: The Rand Corporation, 1972.

Bachman, J. & Johnston, L. Understanding Adolescence. Boston: Allyn and Bacon, 1972.

Bissell, J. The cognitive effects of preschool programs for disadvantaged children. Unpublished doctoral dissertation,. Harvard University Gradeuate School of Education, 1970.

Boyd, J. Project Head Start, Summer 1966: Facilities - Resources of Head Start Centers. Princeton, New Jersey: Educational Testing Service, 1966.

Cicerelli, G., et al. The impact of Head Start: An evaluation of the effects of Head Start on children's cognitive and affective development. (Westinghouse Learning Corporation and Ohio University. Contract B89-4536 with the Office of Economic Opportunity) Washington, D.C. Office of Economic Opportunity, 1969.

Coleman, James, et al. Equality of educational opportunity. U.S. Department of Health, Education, and Welfare, Office of Education, OE-38001, National Center for Educational Statistics. Washington, D.C.: U.S. Government Printing Office, 1966.

Coller, A. & Victor, J. Early Childhood Inventories Project. New York City: Institute for Developmental Studies, New York University School of Education, 1971.

Datta, L. A report on evaluation studies of Project Head Start. Paper presented at the 1969 American Psychological Association Convention, Washington, D.C.

Di Lorenzo, L., et al. Prekindergarten programs for educationally disadvantaged children. Final report Washington, D.C.: U.S. Office of Education, 1969.

Dittman & Kyle Head Start Planned Variation Case Studies. University of Maryland, 1970-71.

Featherstone, Helen: Cognitive Effects of Pre-School Programs on Different Types of Children. Huron Institute, Cambridge Mass. 1973.

Hertzig, et al., "Class and Ethnic Differences in the Responsiveness of Pre-school Children to Cognitive Demands", monograph of SRCD,ee,(1),1968.

Jencks, C., et al. Inequality: A Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, 1972.

Karnes, Merle. Research and development program on preschool disadvantaged children. Final Report, U.S. Department of Health, Education and Welfare, 1969.

Levine, et al. California Preschool Competency Scale Manual. Palo Alto, California: Consulting Psychologists Press, Inc., 1969.

Lukas, Carol and Wollheb, Cynthia: Implementation of Head Start Planned Variation, Huron Institute, Cambridge, Mass. 1973.

Maccoby, E.E., et al. Activity level and intellectual functioning in normal preschool children. Child Development, 1965, 36, p. 761-770.

Maccoby, E.E. & Zellner, M. Experiments in primary education: Aspects of Project Follow-Through. New York: Harcourt Brace, 1970.

McMeekin, R. Costs Analysis of Planned Variation Head Start Models. The Huron Institute, in preparation.

Mosteller, F. & Moynihan, D.P., eds. On Equality of Educational Opportunity. New York: Random House, 1970.

Plowden, Children and Their Primary Schools. Central Advisory Council for Education, London: Her Majesty's Stationery Office, 1967.

Porter, A.G. How Errors of Measurement affect ANOVA, ANCOVA and regression analyses. Paper presented at the 1971 AERA convention.

Rainbow Series. U.S. Department of Health, Education and Welfare, Office of Child Development, 1972.

Rubin, D. Matching to remove bias in observational studies. Technical Report No. 33, December 14, 1970, Department of Statistics, Harvard University.

Shaycroft, Marion. The statistical characteristics of school means. In Flanagan, et al. Studies of the American High School. University of Pittsburgh, 1962.

Smith, Marshall, S. Some Short Term Effects of Project Head Start: A Preliminary Report on the Second Year of Planned Variation--1970-71. Huron Institute, Cambridge, Mass. 1973.

Stanford Research Institute. Evaluation of the National Follow-Through Program 1967-71. Menlo Park, California: draft.

Stanford Research Institute. Implementation of Planned Variation in Head Start. Menlo Park, California: 1972.

Stearns, M.S. Report on preschool programs: The effect of preschool programs on disadvantaged children and their families. Washington, D.C. Office of Child Development, 1971.

Terman, L.M. & Merrill, M.A. Stanford-Binet Intelligence Scale; Manual for the Third Revision Form L-M. Boston: Houghton-Mifflin, 1960.

The Study of Natural Variations in Head Start 1969. Office of Child Development, 1971, draft.

Tukey, J.W. Discussion on Temporal changes in treatment - effect correlation. In Glass (Ed.) Proceedings of the 1971 Invitational Conference on Testing Problems, Princeton, New Jersey, in press.

United States Civil Rights Commission. Racial Isolation in Public Schools. Washington, D.C.: 1967.

Walker, Deoby, Bane, MaryJo, Bryk, and Tony: The Quality of the Head Start Planned Variation Data. Huron Institute, Cambridge, Mass. 1973.

Weikart, D.P. Preschool invervention: A preliminary report of the Perry Preschool Project. Ann Arbor, Michigan: Campus Publications, 1967.

Weikart, D.P. Relationship of curriculum, teaching and learning in preschool education. In J.C. Stanley Preschool Programs for the Disadvantaged. Baltimore: Johns Hopkins University Press, 1972.

White, S., et al. Federal Programs for Young Children: Review and Recommendations. The Huron Institute, 1972.

APPENDIX I

The Curriculum Approaches (Models)

This appendix briefly describes the twelve models used in the Planned Variation study in 1970-71. As we noted earlier, each of the approaches, with the exception of the Enabler model, has been developed and is sponsored by some group of people in a university or private corporation. The descriptions are intended to reflect the goals and expectations of the sponsors rather than to be a critical analysis. As presented, they are idealized descriptions of the twelve treatments. These sponsored approaches were included in Head Start Planned Variations because they were considered to be promising methods for working with disadvantaged children and families and because they were unique in some significant way. Nevertheless, the sponsors share common orientations. All of them seek to develop children's learning abilities. All are convinced of the importance of individual and small group instruction and frequent interchnage between children and concerned adults. All attempt to make learning interesting and relevant to the child's cultural background. All believe that the child's success in learning is inseparable from his self-esteem, motivation, autonomy, and enrironmental support, and all attempt to promote successful development in these domains while fostering academic goals. The sponsors differ among themselves chiefly in the priorities which they assign to these objectives and in the sequences through which they pursue them.

It is important to recognize that the concept of Planned Variation was not intended as a means of finding a single "best" method for educating disadvantaged children. A wide variety of groups of children are included in this study, and a program that is appropriate for some may not be appropriate for others. Some approaches, for example, are primarily concerned with parental involvement and community control, while others place primary emphasis on the curriculum, the teacher, and the classroom. The following paragraphs briefly attempt to capture the emphasis of each model.

EDC Open Education Curriculum
Educational Development Corporation (EDC)

Sponsor Contact: George Hein

EDC has an open classroom approach derived from the British primary school model and theories of child development. It believes that learning is facilitated by active participation in the process. The classroom provides a setting in which there is a range of materials and activities from which the child can choose. Academic skills are developed in a self-directed way through classroom experiences. The role of the teacher is one of leading the child to extend his own work and generally involves working with an individual child or small group.

The Systematic Use of Behavioral Principles Program
(Engelmann-Becker)
University of Oregon

Sponsor Contact: Wesley Becker

The primary focus of the Engelmann-Becker program is on promoting skills and concepts essential to reading, arithmetic and language achievement, with particular emphasis on remedying language deficiencies. The main techniques are programmed materials, structured rapid-fire drills, and positive reinforcements of rewards and praise to encourage desired patterns of behavior. Small study groups of five to ten children are organized by teachers according to ability levels in order to facilitate presentation of patterned learning materials and to elicit verbal responses from children.

The Bank Street College of Education Approach
Bank Street College of Education

Sponsor Contact: Elizabeth Gilkeson

The Bank Street approach emphasizes both learning and social-emotional development of children on the premise that they are intertwined. The teacher functions as a supportive adult whom the child can trust, and teaches by relating and expanding upon each child's response to his experiences. The classroom is viewed as a stable environment and workroom for the child in which he is encouraged to explore, make choices and carry out plans. Academic skills are presented in the context of classroom experiences.

The Behavior Analysis Approach
Support and Development Center for Follow-Through, University of Kansas

Sponsor Contact: Don Bushell

The Behavior Analysis approach has three predominant aspects. First it emphasizes academic and social skills. Individualized programmed materials are the primary teaching mode. Second it makes systematic use of positive reinforcement. A token exchange system is used to support children's learning efforts. Third it employs parents as members of the instructional team as well as behavior modifiers. They receive training and work in the classroom in shifts throughout the year.

Individually Prescribed Instruction and the Primary Education Project (IPI)
Learning Research and Development Center, Univ. of Pittsburgh

Sponsor Contact: Lauren Resnick

The IPI approach provides an individualized program of instruction for each child which teaches him academic skills and concepts in the areas of language, perceptual motor mastery, classification, and reasoning. The materials are sequenced to reflect the natural order in which children acquire key skills and concepts. Diagnostic tests determine each child's strengths and weaknesses and are used by the teacher to prescribe instructional materials appropriate to his needs. Positive reinforcement, both social and concrete, is given continually for success in learning.

The Responsive Environments Corporation Model (REC)
Responsive Environments Corporation

Sponsor Contact: Lori Caudle

The REC model uses specially designed, self-correcting multi-sensory learning materials which strengthen school readiness skills in language and reading. They are designed to teach basic concepts while allowing children to make choices, work independently, and set goals for themselves. Teaching machines in the form of "talking typewriters" and "talking pages" involve children in learning by seeing, tracing, typing, imitating and discriminating among sights and sounds and by recording and listening to their own voices.

The Florida Parent Educator Model
University of Florida

Sponsor Contact: Ira Gordon

The Florida approach is not a specific classroom instructional model but is designed to work directly in the home. It focuses on the parent, believing that the parent is the key agent in a child's development. The major goals of the program are to develop educational competence in the child and to develop an atmosphere in the home which will foster continued growth. An important role is played by paraprofessionals called parent educators. The parent educator spends half-time with the teacher in the classroom and the other half making home visits. The home visit involves bringing tasks into the home and instructing the mother how to teach them to the child.

The Tucson Early Education Model
University of Arizona

Sponsor Contact: Ron Henderson

The Tucson model has a flexible child-oriented curriculum which focuses simultaneously on four areas of development: language competency, intellectual skills, motivational skills and societal skills. Emphasis is placed more on learning to learn skills than on specific content. The content is individually determined by a child's environment and interests. The classroom is arranged in interest centers for small groups. The teacher's role is to work on a one-to-one basis with the child, arrange the classroom setting and encourage interactions between the child, his environment and others. 60043

Responsive Educational Program
Far West Laboratory for Educational Research and Development

Sponsor Contact: Glen Nimnicht

The Responsive Educational model emphasizes self-rewarding learning activities and a structured environment responsive to a child's needs and interests. The model encourages the child to make interrelated discoveries about his social world and physical environment and stresses the importance of the development of a healthy self-concept. The classroom is a controlled environment in which the child is free to explore various learning centers, games and activities. Problem solving and concept formation as well as sensory and perceptual acuity are stressed and the pace of all learning activities is set by the child for himself.

Cognitively Oriented Curriculum
Hi/Scope Educational Foundation

Sponsor Contact: David Weikart

The Cognitively Oriented Curriculum combines Piagetian theory and an open classroom approach. It uses a cognitively oriented curriculum and emphasizes the process of learning rather than particular subject matter. It stresses a child's active involvement in learning activities. The teacher takes an active role. Additionally, home training is seen as part of the program and the teacher suggests tasks for the mother to present to the child at home.

BEST COPY AVAILABLE

The Enabler Model
Office of Child Development

Sponsor Contact: Jenny Klein

The Enabler Model does not involve affiliation with a particular instructional approach. It is build on goals prescribed by each community for itself. The development and implementation of this model are facilitated by the assistance of an OCD consultant who takes a very active role in all aspects of the program. Thus projects with the Enabler Model may differ considerably in the approach and style of their educational tactics, but all share a commitment to high levels of parent participation in policy making, program planning and classroom operation.

APPENDIX II

1. Classroom Information Form: This instrument was used to gather information about the background and family characteristics of every child in the sample. Teachers completed the instrument by gathering information from Head Start application blanks and interviews with parents. A validity study of selected items from a similar form used in 1971-72 yielded encouraging results (see "The Quality of the Data").

2. California Preschool Social Competency Scale: This is a teacher completed rating scale fo 30 items designed to "measure the adequacy of preschool children's interpersonal behavior and the degree to which they assume social responsibility" (Levine, et al., 1969, p.3). An extensive description of the measure is included in "The Quality of the Planned Variation Data". Completion of the scale by teachers suggested to us that among classroom and among site comparisons would be illegitimate. The reason is simply that teachers may consider their own classrooms as the reference group for rating students. Since the compositions of classrooms vary greatly, the ratings may lose comparability when they are taken out of the immediate context of their classroom.

3. Sponsor Ratings of Implementation: This rating form is fully described and analyzed in the report on "Implementation".

4. Head Start Directors Ratings of Implementation: This form is similar to the Sponsor Rating except that it was completed by the Head Start Director. It is discussed in the report on "Implementation".

5. Teacher and Teacher Aide Survey: These forms assess teacher and teacher aide background, teaching experiences and attitudes toward the Planned Variations model. They are extensively analyzed in the "Implementation" report. In this report, we use some items taken from these surveys.

6. Classroom Observation Instrument: This is a broad range objective observation instrument developed at the Stanford Research Institute to assess the degree of implementation of classroom processes and child outcomes in the various programs. Trouble with the coding on the classroom observation tape limited our use of this important instrument. An analysis of some results from it are included in the report on "Implementation" and an extensive analysis of its use in 1971-72 is under preparation by SRI.

7. Basic Child Test Battery: Four tests are included in this battery. The results from these tests are extensively analyzed in this report. Additionally, results from one of the tests, the Caldwell Preschool Inventory, are used in the report on "Cognitive Effects of Preschool Programs on Different Types of Children". Complete descriptions of the tests are in "The Quality of the Planned Variation Data". The four tests are:

a. Caldwell Preschool Inventory. (PSI) The PSI was developed to assess general achievement in preschool in areas deemed necessary for later success in school. Specifically developed for preschool populations, 64 items tap areas of general knowledge, listening and word meanings, listening and comprehension, writing, copying, quantitative skills, and speaking and labeling. Though the test was originally designed to have four factors, factor analyses of our data revealed only one factor which seemed to cut across all areas tapped by the test. Consequently, we simply summed the items to create a score on the test. Internal (KR-20) reliability is roughly .90. By and large, we consider this test a measure of general achievement in preschool.

The scoring procedure for the test is not normed for age and as a consequence, pre-scores on the PSI are highly and positively correlated with the age that the child enters the program. The PSI also correlates roughly 0.50 with the Stanford-Binet, which in turn has a slightly negative correlation with age. The Stanford-Binet IQ score is obtained by dividing a calculated Mental age by chronological age -- the division by age makes the IQ score comparable across ages. The Mental age score taken alone can be thought of as the Binet score uncorrected for age. Mental age on the Binet correlates roughly .75 with the PSI. Assuming both tests have a reliability of .9, we find that the correlation among the "true score" parts of the PSI and the Binet score unadjusted for age is roughly .83*. Though this correlation is far from perfect, it suggests that the two tests are tapping somewhat the same domain.

b. NYU Book 3D. The NYU booklets were designed to measure areas of specific preschool achievement. Book 3D

---

*The sample used for these estimates and other estimates on following pages of this chapter was the same sample used for the correlation matrix on Page 120 of "The Quality of the Planned Variation Data" for estimates of the reliabilities of the two tests.

is designed to tap achievement in pre-math (seven items), pre-science (seven items), and linguistic concepts (five items assessing knowledge of prepositions). Both NYU booklets (3D and 4A) were extracted by SRI from the Early Childhood Inventories developed by A. Collier and J. Victor at the Institute for Developmental Studies at the NYU School of Education. Two scoring systems are used in the analyses in this report. First, a simple summary score obtained by adding together all correctly answered items is used. A factor analysis of the Book 3D suggested that there was only one stable, interpretable factor.* Estimates of internal reliability for the total score are generally in the range of 0.60-0.70. In this report we use 0.65 as a reliability estimate for individual scores. Moreover, the single score seems to have a ceiling problem for some groups of older children on the post-test results. See "The Quality of the Planned Variation Data" for discussions of these issues. Second, a set of scores is obtained by considering the three sub-tests as criterion-referenced measures. Using these measures, we report the percentages of children in various sites

*A factor analysis of Books 3D and 4A together convinced us to keep the tests separated for analytic purposes.

and models for each sub-test who obtain either a perfect score or only one item incorrect at post-test time. We also report the percentages of children in these groups who fail to get more than one item correct on each sub-test.

A score derived from a summing of correct items for Book 3D bears a very strong relationship to the PSI. By and large, different sub-samples of the data reveal correlations of about 0.70 at pre-test time. Adjustment of this correlation for the reliabilities of the two tests (PSI reliability is roughly 0.90 and Book 3D reliability is roughly 0.65) yields a corrected correlation coefficient of roughly 0.95 indicating that the two tests are tapping almost entirely the same domains.

c. NYU Book 4A. This test is designed to tap achievement in three areas: knowledge of alphabet names (nine items); knowledge of numeral names (six items); and knowledge of shape names (three items). The development of scores for this test was similar to the development of scores for the Book 3D. A single summary score is analyzed in this report along with three criterion-referenced measures. With the exception of the third sub-test we follow the

same rules for creating our criteria, as we did for Book 3D. In the third sub-test, we required that the student answer all three questions correctly to meet the criterion. The single score on Book 4A has an internal reliability of roughly 0.65 for the pre-test. To some extent this reliability is reduced by a minor floor problem in the Fall testing. For all groups the Book 4A scores were positively skewed in the Fall and more normally distributed in the Spring. Pre-test scores for Book 4A and the PSI correlate roughly 0.45-0.50, with the Book 3D the correlations are roughly 0.40-0.45 and with the Stanford-Binet, the correlations are roughly 0.40. Overall, then, though the Book 4A is assessing somewhat similar areas as the PSI, Book 3D and the Stanford-Binet, there is considerable residual unique variance associated with the test.

d. Motor Inhibition Test. This test was developed by Hagen and Degerman (see Maccoby et al., 1965) to measure a child's ability to inhibit movement when the task demands it. Three tasks are used to assess inhibition, the Draw a Line slowly task, the Walk slowly task, and the Pull Truck slowly task. Four preliminary items assess the child's understanding of the concepts of slow and fast. A substantial propor-

tion of the sample of children in this study (over 50%) failed to answer two or more of the four pre-test items correctly, in either the Fall or Spring, indicating that these children did not understand the two concepts. The scores on the Motor Inhibition test were not analyzed for these children. Analyses of the three sub-tests indicated that the first two tasks yielded scores that correlated roughly 0.46. Correlations of the first two tasks with the third task were roughly 0.24. The low correlations with the third sub-task indicated to us that it was either unreliable or was measuring something other than the other two sub-tasks. Consequently, we formed a measure of the Motor Inhibition by summing the amounts of time in seconds taken to complete the first two sub-tasks. Following Maccoby's lead and an inspection of the data, the log of this score was then taken. The log transformation removed the strong positive skewness from the new scores. This final score correlates in the 0.30 to 0.40 range with the NYU 3D and PSI and in the 0.15-0.20 range with the Book 4A and the Stanford-Binet.

8. Ethnic Heritage Test: Two tests were actually used here. The Ethnic Identity Questionnaire (EIQ) was developed by Manuel Ramirez III at the University of California, Riverside, to investigate the ethnic identity of Mexican-American children and the Children's Cultural Awareness Scale (CCAS) was developed by Edward J. Barnes at the University of Pittsburgh to explore the cultural awareness of Black children in the Head Start Planned Variation Study. Scores from neither test are used in this report.

9. Stanford-Binet: The Stanford-Binet Intelligence Scale is a well-known measure of "general intelligence". The 1960 revision was used in this study. A single measure of IQ is used in this report. After extensive checking for matched pre- and post- birthdates and valid items, the score was calculated by dividing a child's Mental Age derived from the test by his chronological age in months and then corrected for age-related fluctuations in variance using the revised Pinneau tables (see Terman and Merrill, 1960).

10. 8-Block Sort Task Test: The Eight Block Sort Task is a measure of maternal teaching style and interaction styles between mother and child. The score used in this report ranges from 0-8 points and indicates the success of the mother in teaching the sorting tasks to the child. (See "The Quality of the Planned Variations Data" for an extensive discussion of this measure.)