DOCUMENT RESUME

ED 045 196 PS 003 678

AUTHOR Miller, Louise B.; Dyer, Jean L.
TITLE Experimental Variation of Head Start Curricula: A Comparison of Current Approaches. Annual Progress Report, June 1, 1969 - May 31, 1970.
INSTITUTION Louisville Univ., Ky. Dept. of Psychology.
SPONS AGENCY Office of Economic Opportunity, Washington, D.C.
PUB DATE 31 May 70
NOTE 119p.

EDRS PRICE EDRS Price MF-$0.50 HC-$6.05
DESCRIPTORS Comparative Analysis, Educational Philosophy, Effective Teaching, Grouping (Instructional Purposes), Preschool Children, *Preschool Programs, Program Content, *Program Descriptions, *Program Effectiveness, Teacher Characteristics, Teacher Education, Teaching Techniques, *Testing, Tests
IDENTIFIERS Bereiter Engelmann, Montessori, Project Head Start

ABSTRACT

Two major questions were raised in regard to four types of preschool programs. (1) Do programs differ in actual operation as well as descriptively? (2) Do programs have significantly different effects on children? Analysis of data obtained on samples of four classrooms in each of three programs (Bereiter-Engelmann, DARCEE, Traditional) and two classrooms in the fourth program (Montessori), showed clear differences. Despite within-program teacher differences on variables assessed by monitoring procedures (observation, television), results provide no evidence that the teachers' (N=14) characteristics were a source of difference among programs. Results also indicate that a brief four to eight week teacher training program supplemented by visits from consultants is adequate for identifiable program implementation. The four programs did have significantly different effects on children's cognitive, social, and/or motivational development. The immediate effects of Bereiter-Engelmann and DARCEE were statistically significant in academic and motivational development. The effects of Bereiter-Engelmann were largely confined to cognitive and academic areas. The effects of DARCEE were more diffuse and most evident in the areas of motivation and attitudes. Sex differences occurred. In general, results indicate that the immediate impact was superior for the two most didactic programs. (WY)

"EXPERIMENTAL VARIATION OF HEAD START CURRICULA: A COMPARISON OF CURRENT APPROACHES"

Research Grant #CG 8199

from

Office of Economic Opportunity

ANNUAL PROGRESS REPORT

June 1, 1969 - May 31, 1970

Louise B. Miller, Ph.D.
Project Director

Jean L. Dyer, Ph.D.
Research Associate

CHILD DEVELOPMENT LABORATORY
Psychology Department
University of Louisville
Louisville, Kentucky

CONTENTS

		Page
I.	INTRODUCTION	1
II.	EXPERIMENTAL PLAN	4
III.	METHOD	
	A. Design	5
	B. Treatment Descriptions	8
	Bereiter-Engelmann	8
	DARCEE	12
	Montessori	14
	Traditional	17
	C. Treatment Implementation	20
	Teacher Training	20
	Classroom Operations	20
	Parent Involvement	21
	D. Assessing Treatments	22
	Verification of Intact Treatments	22
	Analysis of Treatment Dimensions	22
	Selection of Dimensions	23
	Construction of Monitoring Procedures	24
	E. Assessing Teachers	32
	F. Assessing Treatment Effects	33
	Selection of Instruments	33
	Recruitment and Training of Testers	36
	Procedural Controls	37
	Rater Comparisons	38

IV. PREDICTIONS

A. Treatment Dimensions 39

B. Treatment Effects 40

V. RESULTS

A. Design 41

B. Profiles 44

C. Treatment Characteristics 44

Verification of Intact Treatments 44

Analysis of Treatment Dimensions 49

Teaching Techniques 49

Grouping Patterns 55

Intercorrelations 57

Summary 61

D. Teacher Characteristics 62

Personality 62

Intelligence 62

Attitudes Toward Teaching 62

Agreement with Program Philosophy 62

Summary 64

E. Treatment Effects 64

Across Programs 64

Method of Analysis 64

Cognitive Measures 67

Motivational and Social Measures 75

Perceptual Measures 94

Intercorrelations 98

Summary 103

Within Programs 104

Class Differences 104

DARCEE Home Visitor Study ... 104

F. Relations between Treatment Dimensions and Treatment Effects 104

VI. SUMMARY AND DISCUSSION 106

VII. REFERENCES 107

LIST OF FIGURES

No.		Page
1	Design of Experiment Replication by Geographical Areas	6
2	Video Tally Sheet - Giving	25
3	Video Tally Sheet - Asking	26
4	Video Tally Sheet - Miscellaneous	27
5	In-Class Tally Sheet	29
6	Standard Score (Z) Profiles for In-Class Monitoring on Teachers and Children - Cell/Tally and Cell/Row Measures	45
7	Standard Score (Z) Profiles for Video-Tape Monitoring on Teachers	46
8	Standard Score (Z) Profiles for Treatment Effects	47
9	Fall-Spring Means on Stanford-Binet for Programs and Controls	69
10	Cumulative Distribution of Stanford-Binet IQs	71
11	Fall-Spring Means on Preschool Inventory for Programs and Controls.....	72
12	Fall-Spring Means on Quick Picture Vocabulary for Programs and Controls	73
13	Male-Female Means on Stanford-Binet for Programs and Controls	76
14	Male-Female Means on Preschool Inventory for Bereiter-Engelmann, DARCEE and Traditional Programs	77
15	Fall-Spring Means for Males and Females on Preschool Inventory for Programs and Controls	78
16	Fall-Spring Means on Curiosity-Verbal for Programs and Controls	80
17	Fall-Spring Means on Curiosity-Activity for Programs and Controls	81

		Page
18	Fall-Spring Means on Persistence (Replacement Puzzle) for Programs and Controls	82
19	Fall-Spring Means on Resistance to Distraction (Replacement Puzzle) for Programs and Controls	83
20	Fall-Spring Means on Inventiveness (Dog and Bone) for Programs and Controls	84
21	Fall-Spring Means for Males and Females on Inventiveness (Dog and Bone) for Programs and Controls	85
22	Spring Means on Stanford-Binet Face Sheet Ratings by Binet and Preschool Inventory Testers for Programs and Controls	91
23	Male-Female Means on Combined Factors of the Stanford-Binet Face Sheet Ratings by Preschool Inventory Testers	92
24	Fall-Spring Means on Embedded Figures for Programs and Controls	95

LIST OF TABLES

No.		Page
1	Demographic Information on Program and Control Subjects	42
2	Fall-Spring Test Intervals and Number of Testers Used	43
3	Consultants' Ratings of Programs	48
4	Means on In-Class Monitoring for Teachers' and Children's Behavior	50
5	Means for Video-Tape Monitoring of Teachers' Behavior	53
6	Grouping Patterns and Total Activity	56
7	Correlations Among In-Class Monitoring Variables	58
8	Teachers' Agreement with Statements from Program References	63
9	Consultants' Agreement with Statements from Program References	65
10	Fall-Spring Means on Main Test Battery for Programs and Controls	68
11	Means on Additional Spring Tests for Programs	74
12	Fall-Spring Means on Behavior Inventory Ratings by Teachers and Aides	86
13	Teacher-Aide Correlations on Behavior Inventory	88
14	Means on Stanford-Binet Face Sheet Ratings by Binet and Preschool Inventory Testers for Programs and Controls	90
15	Binet-PSI Tester Correlations on Stanford-Binet Face Sheet	93
16	Means for Error Score on Wepman and Correct Responses on CADI	97

No.		Page
17	Fall-Spring Correlations Among the Dependent Variables for Subjects in the Experimental Programs	99
18	Spring Correlations Among the Dependent Variables for Males and Females in the Experimental Programs	100

FOREWORD

The Annual Progress Report for the year 1969-70 summarizes and condenses material from all previous reports including the first Annual Report for the year 1968-69, Progress Reports 4 and 5, and the Proposal for Continuation of the Grant. Thus it should not be necessary to refer back to previous reports for information on the first two years of the project.

This Annual Report also contains the following new information obtained since submission of the 5th Progress Report in February:

(1) Results from analysis of variance by sex for experimentals and controls

(2) Results of the Stanford-Binet Face Sheet ratings of children by both Binet testers and Preschool Inventory testers

(3) Correlations between the two groups of testers on the Binet Face Sheet

(4) Comparisons on all dependent variables between children who were visited regularly by Home Visitors in the DARCEE program with results for children who were never visited

(5) Comparisons between controls who were obtained from waiting lists for Head Start and controls who were recruited in neighborhoods

(6) Results from the Behavior Inventory ratings by aides in the classroom

(7) Correlations between teachers and aides on Behavior Inventory ratings

(8) Results of statistical analyses of demographic data

(9) Results from a study comparing the Wepman Auditory Discrimination Test and the California Auditory Discrimination Index

(10) Results from analysis of variance of the video-tape monitoring in prekindergarten, indicating the significant sources of variance such as reinforcement

(11) Intercorrelation matrices for in-class monitoring variables

(12) Intercorrelation matrices for all dependent variables

(13) Intercorrelation matrices separately by sex for dependent variables

(14) Results from statistical analysis of teacher personality (16 PF), biographical data on teachers, teacher attitudes toward teaching (Neill's Questionnaire), and teacher IQ (Peabody Picture Vocabulary Test).

In addition to analyses of data from the prekindergarten comparisons, the following activities were completed during the 1969-70 year:

(1) Testing at the end of the kindergarten year of all experimentals and controls. Included in this group are 29 children not previously tested who were added to the control group for the purpose of various comparisons.

(2) Testing of 48 four-year-old middle-class controls. The results from this group will be combined with results obtained in 1968-69 on experimentals and lower-class controls and data will be re-analyzed.

(3) Initiation of work on the production of a film report of the research.

(4) Video-tape monitoring of kindergarten classes containing children from the experimental programs and controls. Tapes were made on three occasions in both Follow-Through and regular kindergarten classes in 10 classes.

INTRODUCTION

As early as 1967 there was a fair amount of evidence that a variety of intervention programs at the preschool level could produce gains on various measures of intellectual ability. In a review of such programs up to that time, Hodges and Spicker (1967) concluded that "intervention programs especially designed to remedy cognitive deficits during the preschool years and to prevent progressive school failure during the later school years have been relatively effective to date." Bellar (1969) has provided recent evidence that children who had special kindergarten or prekindergarten programs were receiving better grades as late as the third grade in more than 50 different schools. Gray and Klaus (1969) report evidence of the superiority of their two experimental groups over controls at the end of a 6-year period. The difference is modest but statistically significant.

Unfortunately, there is also a body of evidence on what has become known as the "fade-out phenomenon". That is, the gains made by children in special preschool programs are sometimes not maintained for more than a year or, at the most, two years, after such special programs have been completed. One of the most recent studies is that of Karnes (1968) who reports that dramatic increases in academic programs had faded by the end of the first grade. The fading of results which a number of investigators have reported led Jensen in his review (1969) to question the nature of intervention effects--that is, are they "hot-house" or "fertilizer"? Restating the problem, he says, "There remains the question of the extent to which specific (early) learning affects cognitive structures which normally do not emerge until 6 or 7 years of age and whether induced gains at an early level of mental development show appreciable 'transfer' to later stages. It is hoped that investigators can keep sufficient track of children in preschool programs to permit a later follow-up which can answer these questions." (p. 106).

This question of what is really changed as a result of successful intervention programs cannot be lightly dismissed, for it bears on the fundamental nature of the development of cognitive functioning. If it is the case, as suggested by Skinner's methods, Gagné (1963), and others, that intellectual ability develops cumulatively, each advance being dependent upon strategies and skills developed earlier, then acceleration in the more basic skills should produce a permanent advantage (assuming that the "basic" skills at each level can be identified).

On the other hand, it is a reasonable hypothesis that the full development of potential in cognitive ability is more related to motivations and attitudes developed in the preschool years than to specific skills at an early age. This notion is supported by several lines of evidence, such as effectiveness of work with parents (Miller, 1967) and the more "middle-class" home environment found for children who do maintain their gains (Kirk, 1958).

These two hypotheses are not necessarily mutually exclusive, of course, since cognitive development as measured by available tests may depend on a combination of skills and attitudes. This latter position is the basis for the DARCEE (Early Intervention) Program developed by

Klaus and Gray at the George Peabody University.

Of equal importance, both in terms of the development of cognitive theory and in a practical sense, is the question, "Which aspects of successful programs are responsible for their effects?" Programs usually make many modifications simultaneously. In the Hodges and Spicker review (1967), previously referred to, the authors also commented, "No one approach at this time appears to be more effective than any other". Kounin (1969) suggests that the ecology of the classroom is more important than teacher or child personality. Recently, Weikart (1969), on the basis of "no difference" results for three programs which he studied (all of which raised IQ's substantially), offered the hypothesis that some of the common elements in special programs which account for their success are probably commitment and enthusiasm, use of a specific model, and the organization of the intervention effort, rather than the nature of the specific program.

The problem of the relative effectiveness of programs requires for its solution the identification of the components or dimensions of programs in their actual operation. Gordon (1969) has effectively stated this point of view by saying, "We need to engage in a very systematic observation: the kind of monitoring or quality control, if you will, for taking samples over time of what is actually transpiring to see whether or not they are doing what the model says they ought to be doing and whether or not in reality the models they say they hold really differ when they become operational...on what dimensions are they alike and on what dimensions are they different? Only when we begin to address ourselves carefully to that kind of question will we have the kind of research and evaluation that tells us directions for new programs." (Seminar #2, p. 38).

Although a number of methods have been used to monitor classroom activity, none of these provide the kind of data that would link the variables known to be important in learning with the dimensions of specific programs.[1]

Rashid (1969) concludes that there is a sharp break between the professional literature describing practices in preschool programs and the literature describing practices in the primary grades, in that most of the research at the preschool level has been concerned with the teacher's influence on behavior in the general domain of personal and social development rather than teacher competence or effectiveness.

Fortunately, a number of important issues regarding the variables which affect intellectual development are brought into sharp focus by the philosophies and techniques of several of the most successful programs designed to modify such development during the preschool years. A few of these variables are: language (verbal instruction as a technique by teachers and practice in language use by children), imitation and its converse, modeling by teachers, role-playing, reinforcement, manipulation

[1]These have been reviewed by Sears and Dowley (1963) and by Biddle and Ellena (1964).

of materials and sensorial stimulation. If programs could be ordered along these dimensions in terms of actual classroom activity, as should be possible if they are correctly implemented, then the effects of the various components might be assessable.

A number of troublesome methodological questions regarding intervention efforts also remain unanswered. For example, programs have usually been evaluated either by program developers themselves or by researchers interested in particular kinds of intervention programs. As late as 1969 Jensen pointed out that "A further step in proving the effectiveness of a particular program is to demonstrate that it can be applied with comparable success by other individuals in other schools, and if it is to be practicable on a large scale, to determine if it works in the hands of somewhat less inspired and less dedicated practitioners than the few who originated it or first put it into practice on a small scale". (1969, p. 102). This problem of the Hawthorne effect is clearly described by Sprigle, Van de Riet and Van de Riet (1967) in comments on their own study.

A related problem concerns the teacher variable. In many evaluations different programs have been represented by single classes, a design which of necessity confounds teacher effectiveness with program effectiveness.

Finally, there is the question of the characteristics of the disadvantaged population with whom the program is used. Variables are likely to be differentially effective as a function of many group and individual differences, including among the latter the child's level of development at the time the program is introduced.

It is clear that a number of very important questions regarding preschool education remain at this point unanswered. Further, it appears likely that the answers to some of these questions would not only lead to improvement in the quality of preschool education but would also provide information relevant to more basic, theoretical problems in the area of early development, particularly with respect to learning and its relation to cognitive development.

II. EXPERIMENTAL PLAN

This study was designed to compare the dimensions and the effects of four preschool programs for disadvantaged children.

These four programs were: Bereiter-Engelmann which emphasizes acquisition of linguistic and numerical skills by use of verbal instruction, imitation, and reinforcement, and de-emphasizes sensorial stimulation and manipulation; DARCEE which emphasizes, in addition to verbal and conceptual skills, the acquisition of attitudes and motives related to learning, using verbalization, reinforcement, manipulation of materials, and imitation; Montessori which emphasizes development of persistence, independence, and self-discipline, in addition to conceptual skills, using sensorial stimulation, manipulation of materials, and self-selection, and de-emphasizes reinforcement and verbalization; and Traditional (official Head Start Program) which emphasizes development in social and emotional areas, language skills and curiosity, using manipulation of materials, sensorial stimulation, role-playing, and self-selection, and de-emphasizes verbal instruction and reinforcement.

Of major importance in the study was the attempt to overcome methodological weaknesses common to curricular comparisons. Thus, the study was designed to eliminate the confounding of teacher and program by providing an adequate sample of teachers, to provide several samples from the target population, and to incorporate two control groups--a non-preschool group similar to the experimental sample and a middle-class group in a private preschool. The design as implemented provided for four replications of comparisons among three of the programs, and two replications of comparisons among all four. In addition, much emphasis was placed on determining the dimensions of treatments and in selection of a broad range of instruments to assess treatment effects.

Teachers were trained by program developers or their representatives for 4-8 weeks prior to the experiment. Fourteen classes were conducted during the 1968-69 school year--two Montessori classes and four classes in each of the other program styles. Four-year-olds, randomly assigned within schools to Head Start classes, were tested in the fall after about 8 weeks of school and again in the spring at the end of the school year. Nine instruments designed to assess gains in cognitive, motivational, social and perceptual development were used. Five additional tests were administered at the end of the year to a smaller sample, primarily to assess specific skill-learning. Monitoring was done in-class and by video-taping five times during the year to assess treatment dimensions for both children and teachers.

Within the DARCEE program, two Home Visitors were used, each visiting regularly with half the parents in two classes.

All classes were conducted as a part of the ongoing Head Start Program, with facilities and ancillary services provided by the Community Action Program through its Delegate Agency, the Louisville Public Schools. Curriculum was determined and supervised by the research staff.

III. METHOD

A. DESIGN

Subjects

The total sample consisted of 296 four-year-old children. There were 214 experimental subjects, 98 males and 116 females, enrolled in Head Start classes in poverty areas in Louisville, Kentucky. Ninety-two percent of these children were Negro.

Two control groups were also tested. One, a low-SES group (C-LC), consisted of 34 children of the same age and from the same neighborhoods as the experimental sample--18 were male, 16 female. These controls were not attending a preschool or day care center. The most suitable pool from which to draw control subjects was the waiting list for Head Start classes. These children should have characteristics similar to those of the children enrolled in experimental classes. Consequently, these waiting lists were exploited fully, even though some loss was anticipated due to children going into experimental classes to replace dropouts. Of the 34 tested, 21 were on waiting lists and 13 were suggested by teachers and principals in the schools where experimental classes were located. In many cases they were younger brothers and sisters of children in elementary school.

The second control group consisted of middle-class four-year-olds (C-MC) attending three classes in a private preschool. A total of 48 children--25 male, 23 female--were tested. This control group was obtained during the year following the experimental comparisons.

Experimental Replication

Design of the experiment and placement of classes is shown in Figure 1. The four experimental replications were placed in four "target areas" of the city. These geographical areas as defined by the Community Action Commission were larger than neighborhoods, but smaller than census tracts. Descriptions of the areas based on the 1960 census indicated that they differed in respect to unemployment, average income, and a number of other factors. Thus it was important to assess sample characteristics and to balance classroom facilities across programs. Since the experiment was being conducted as a part of the regular Head Start Program, it was also desirable to provide both experimental and non-experimental classes in the same schools.

The four target areas were designated California, Jackson, Park-DuValle and Russell. Since the 4-program comparison could be replicated in only two areas with the two Montessori teachers, the two largest areas, Park-DuValle and Russell, were selected for this purpose. All four areas contained replications of the 3-program comparison. Classroom facilities in the Russell area were in general inadequate. All four programs were located outside of school

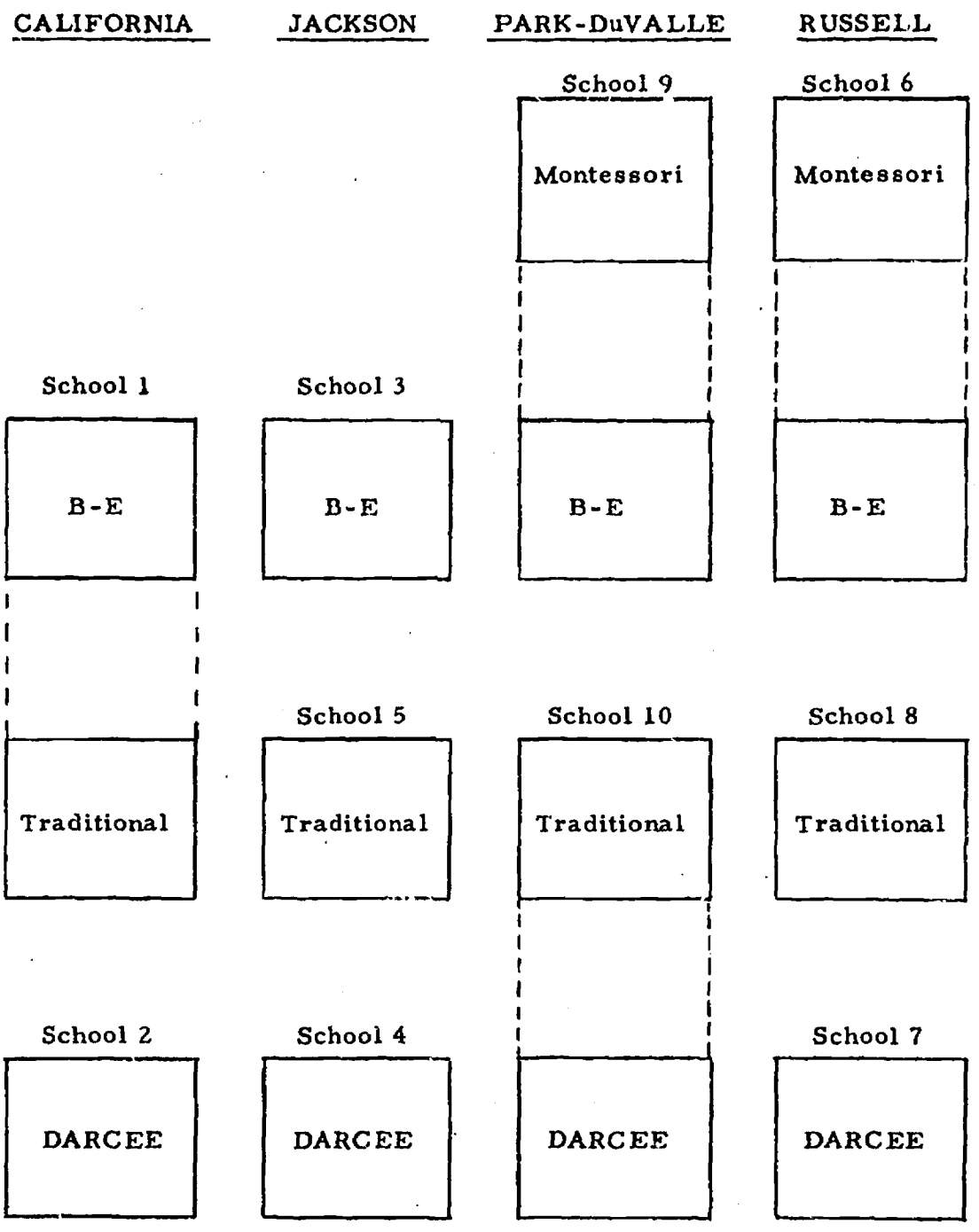

Fig. 1. Design of Experiment. Replication by Geographical Areas.

buildings, three in churches and one in a small and very old portable. In the other three areas, facilities were in satisfactory school classrooms.

Random Assignment of Subjects

To some extent the distribution of classes constituting the various programs into different geographical areas insured that the combined program samples would consist of similar subjects. But it was also desirable that children who attended experimental classes would constitute a random sample of those who registered for Head Start.

In the strict sense, "random" assignment of subjects would be accomplished by obtaining the names of all children eligible for Head Start and, assigning each child to one of the 14 classes or to a control group by using a table of random numbers. This would not have been possible with a sample of approximately 250, since it would have forced many children to cross the city to attend schools outside their neighborhoods.

It was possible, however, to arrange for assignment of registrants in each school on a random basis. All schools contained at least two Head Start classes. In one school, both classes were experimental; however, in the remaining schools both experimental and non-experimental classes were available for distribution of subjects. Registration forms were filled out on the same day in all schools, including the ten containing experimental classes. The parents' signature on the form gave permission for children to be placed in experimental classes should they happen to be selected.

When all forms were turned in, the forms were divided on the basis of sex to insure a balance in each class. These piles were then shuffled and distributed into classes, experimental or non-experimental.

As can be seen by reference to Figure 1, even in the event that this procedure had not been followed, it would have been impossible for particular programs to be affected by selective assignments, since quite different options were available in the various schools.

B. TREATMENT DESCRIPTIONS

Verbal descriptions of the four programs compared in this study are based on observations of representative classes, observation of training programs, and source materials of the following kinds: (1) publications by program developers, (2) publications recommended to teachers as required reading or required materials in training programs, (3) lectures and workshops during pre-service and in-service training, (4) personal communication with program developers, consultants, and others involved in training programs. Complete consistency is not to be found among these sources nor within them. The most difficult program to describe is the Traditional. In this case much weight was given to the Rainbow Series published for Head Start (1965) and to Hymes (1968).

Bereiter-Engelmann

1. Philosophy and Goals

The Bereiter-Engelmann program is remedial, emphasizing acquisition of the tools of academic learning - verbal and numerical symbols. The approach is pragmatic, preparatory, and selective, focusing on academic deficiencies and the necessity for acceleration. Increases in global IQ are expected, improvement in ability to handle numerical concepts and operations, and the ability to understand and use language - e.g., speak in complete sentences, and understand negation, plurality, logical inclusion and exclusion. Increased self-confidence and satisfaction with self are seen as by-products of success in academic areas.

The target population is described as consisting of children from environments characterized by disorder, lack of discipline, and infrequent reward for intellectual effort. These children, who lag behind their middle-class peers, lack motivation to learn, do not value verbal praise from adults, and are deficient in a language which is adequate for academic use. Many of them manifest the "great word syndrome" which is the use of phrases as units of speech - e.g., "dabidaw" for "that's a big dog". This chunking of units larger than single words makes it difficult for the disadvantaged child to handle words, one evidence for this being the inability to reverse the order of words in a short sentence. The authors do not believe that there are peculiar emotional needs or problems associated with cultural deprivation.

2. Curriculum Content and Organization

The curriculum in the Bereiter-Engelmann program is organized into three areas: reading, language and arithmetic. (1) The reading program is essentially phonetic. Children learn to

recognize and pronounce consonants, vowels, and blends. The short and long vowels are identified with appropriate marks. They then learn to combine these sounds, pronouncing them together to form "words", some of which are meaningful, some are not. For example, children might learn to pronounce the short "a", the "t", and a number of initial consonants such as the hard "g", "s", and "r". They would then read "gat", "sat", "rat". Not all sounds or letters are taught initially; thus, depending upon the stage of the reading program which a given child has reached, he might or might not know all of his letters. (2) The arithmetic program is built around counting operations such as counting towards a number, counting from a number, counting towards a number from a number and counting backwards, and relates these to addition, algebraic addition, subtraction, and algebraic subtraction. Another important concept throughout the program is equality. (3) The language program is oriented toward the structural and logical components of language, emphasizing, for example, recognition of negation, compounds and plurals, speaking in complete sentences, and the use of words as basic units. The first-year language program begins with labeling common objects and proceeds to polars, prepositions, if-then statements, same-different, before-after, pronouns, verb tense and function words. Thus, when a child has successfully completed the first-year program, he has the insturctional language needed to function successfully in an average classroom.

Task analysis and sequencing are extremely important and are provided for the teacher in her instructional workbooks. Each of the three academic areas is programmed for the teacher. Each step should be thoroughly mastered before the group is allowed to proceed to the next step.

Printed material, usually accompanied by drawings, is constantly used in conjunction with verbal instruction, and teachers also make frequent use of small chalkboards on which numbers, letters, and blends are written.

Children are grouped roughly according to ability, ideally into three groups of approximately five each, with three teachers, one for each academic area. Each teacher specializes in one of the areas and teaches it to all three groups. Group instruction utilizes a procedure called "patterned drill" which consists of modeling by the teacher followed by elicitation of unison responding on the part of the children. There is also frequent but brief attention to individual children during group instruction. Very slow learners may be given individual instruction at other times. The pace is rapid but extremely repetitive. Group drill in each of the three academic areas lasts for 20 minutes, with groups of children rotating among teachers. Thus each child receives a total of one hour of patterned drill in a given day. There is also some whole-group activity, typically music, and at other times the child may select among a limited number of activities and materials.

The teacher is in control of task selection and should pace the program so that children in all groups will experience success, but at the same time be challenged by new tasks. (Cf. Montessori). The child does not make his own selection or influence the curriculum by his interests. What he is ready for is determined by his responses, which should be observed and evaluated by the teacher. The atmosphere in the Bereiter-Engelmann classroom is "business-like and task-oriented". It should not be grim or repressive, and children should be lively, not apathetic. During patterned drill, however, inattentiveness is not allowed, nor can the children leave.

3. Methods and Techniques

Virtually the entire Bereiter-Engelmann program is linguistic. Drills are carried on by modeling of language and elicitation of linguistic production by the children. Conversation is not emphasized.

Reinforcement is used continually in a contingent manner as a means of shaping and sustaining desired responses and patterns. It is considered essential to provide the child with specific information regarding the correctness of his responses, and therefore both positive and negative reinforcements are used, primarily as knowledge-of-results. The initial use of primary reinforcement (usually raisins) is advocated but is always accompanied by verbal and social reinforcement until these become sufficiently powerful in themselves that primary reward can be eliminated. This shaping procedure should culminate in self-praise or pride in accomplishment.

Pure sensory stimulation plays virtually not role in the Bereiter-Engelmann program. Although the authors recognize that severe sensory deprivation can be damaging, especially in the first few months of life, they maintain that most disadvantaged children receive as much sensory stimulation as is necessary. If children are allowed unlimited choice of activities in an object-rich environment, they will flit from one thing to another, leaving an item as soon as its sensory qualities become familiar.

Practice and the formation of response habits are greatly stressed. The programs are structured to insure repetition in the form of frequent and systematic review.

Manipulation of concrete materials does not play any significant role in the Bereiter-Engelmann program.

Imitation of the teacher is primarily what children are doing during patterned drill, and is an important technique. Imitation as used in this program, is closer to the meaning of matched-dependent behavior as defined by Miller and Dollard (1941) than to the Freudian notion of indentification.

Friendly competition and calling of attention to successful performance by others is a common technique in this program in contrast to Montessori and Traditional. "Let's see who can get it right first", and "Let's see if we can all do it as well as Jimmy did", are typical of techniques used by Bereiter-Engelmann teachers. There is also competition with the teacher, carefully structured by her so that the children frequently win.

DARCEE

1. Philosophy and Goals

Two major goals characterize the DARCEE program: (1) remediation of linguistic and conceptual deficiencies, and (2) development of a number of attitudes which are related to academic achievement. The orientation is primarily remedial and focused on intermediate goals for all children. However, there is a heavy emphasis on working with parents in an attempt to extend curriculum goals into the home, and in this sense the program is less specifically preparatory than Bereiter-Engelmann. Special attention is given to the development of motivation to achieve, persistence in tasks, resistance to distraction and delay of gratification. Academically, in addition to development of linguistic skills, emphasis is also placed on classification, information about the world, visual and auditory discrimination, and the ability to handle concepts such as time and space. Progress is expected with respect to behavior appropirate to the school situation--sitting still, paying attention, following directions, using verbal rather than physical persuasion and having respect for persons and property.

The DARCEE program recognizes the same kinds of deficiencies in disadvantaged children as described by Bereiter-Engelmann, specifically lack of academic motivation and linguistic deficiencies. In this program neither of these deficiencies has priority over the other since they are seen as part of the same problem and it is considered as necessary to build in appropriate attitudes as to teach concepts.

2. Curriculum Content and Organization

The skill development portion of the DARCEE curriculum is organized around three processes: (1) Input, (2) Association processes, and (3) Output. In other words, the curriculum is designed to help children perceive, decode, and encode stimuli through all sensory channels, to develop skills of association, classification, and sequencing, and to develop the skills necessary for effective verbal communication and expression of thought patterns. Within these three process categories each skill is organized along two dimensions: (1) from a gross elementary level of discrimination to a more specific and complex level, and (2) from concrete to abstract. With respect to content,

the curriculum is organized around units. The first unit, for example, is about the child, then comes a unit about pets, one about seasons, etc.

Sequencing is extremely important in the DARCEE program. For example, in the area of auditory discrimination gross sound identification (loud-soft, high-low) is introduced first. These concepts are then refined to the comparative and finally the superlative distinction. Instruction begins with concrete objects present and moves toward identification of sound alone. Complexity of verbal directions is increased by demanding more precise responses and multiplying the number of directions in a specified sequential order. At a later time whole-word discrimination is introduced and eventually words are made more and more similar until only initial consonants differ. According to Cupp (1967), when one-letter sound distinctions can be discriminated, the child is ready for work in sound-letter association in direct preparation for reading.

Children are grouped in the DARCEE program as they are in the Bereiter-Engelmann program according to initial level of performance. Similarly there should be three adults in a classroom, one for each group of about five children. Throughout the day the children work in groups, all members of the group doing the same thing.

The teacher is a very active agent in the learning process with a role comparable to that of teachers in the Bereiter-Engelmann program. She determines what activities take place and when, and directs these in the way she believes will provide the greatest amount of learning. Children in the DARCEE program are sometimes allowed to play, and given some choice of activities but even these free-play activities are directed towards learning something.

Classroom atmosphere should be both quiet and orderly. Children are required to sit straight in their chairs and they leave only when told to do so. Speaking in loud voices is not permitted. Children line up whenever they are moving from one room to another. The teacher's relationship to the child should be warm but firm. Her main role is teaching, not providing emotional support or being a substitute mother.

3. Methods and Techniques

Language occupies a more prominent place in the DARCEE program than in any other of the four with the exception of Bereiter-Engelmann. The difference between these two perhaps lies less in the degree of emphasis on language than in the method used to promote its development. In addition to formal instruction with linguistic material conversation between teacher and children is encouraged in DARCEE particularly during small group times and

at snack and meal times. This is a primary technique in developing expressive skills and utilization of information.

Heavy emphasis is placed on the importance of contingent reinforcement. There is less emphasis on the correction of errors and more on positive reinforcement for correct behavior, but it is clear that in either case the child should get immediate feedback regarding his responses. Candy reinforcement may be used in the early stages of the program but again this is always accompanied by verbal and social reinforcement and should be eliminated as soon as the latter becomes effective.

Sensory stimulation per se is not emphasized in the DARCEE program but the materials do provide a much wider range of stimuli with which the children interact than is the case in Bereiter-Engelmann. The basic five--beads, parquetry blocks, puzzles, peg boards, and counting cubes--are all concrete objects which offer opportunities for stimulation in various modalities, as well as manipulation. In addition, sensory discrimination, in all modalities, is a formal part of the program.

The role of practice is not greatly stressed in materials from the DARCEE program but the use of the same materials in a variety of ways insures a certain amount of repetition. In addition, the sequencing of the program within units assures practice until a certain level is mastered.

Manipulation of concrete materials is a very prominent aspect of the DARCEE program. Children spend much time manipulating beads, blocks, picture cards, etc.

Imitation in the matched-dependent behavior sense is also a very important part of the DARCEE program. Teachers frequently make towers, designs, etc., which the children are asked to copy.

Competition is not stressed, but neither is it forbidden. Teachers utilize it indirectly in calling attention to appropriate behavior on the part of individual children in giving lavish praise to the children in a way which suggests that their behavior is to be imitated.

Montessori

1. Philosophy and Goals

The Montessori program is intended to be an educational philosophy extending from preschool through the twelfth grade. Individual children may be advanced in some areas and retarded in others. Thus it is not possible to set standards at an intermediate level which are to be expected of all children. The program is not preparatory in the academic sense but is

focused rather on long-term developmental processes. There is, however, heavy emphasis on cognitive development. Academic materials are designed to teach concepts such as weight, length, volume, number, letters, etc.

The goals of a Montessori preschool program fall into four general categories: (1) development of the senses, ability to discriminate, identify, and match, (2) conceptual development, including mathematical concepts, size, weight, volume, etc., (3) competence in daily activities of the kind involving housekeeping and personal care, and (4) what might be called character development--the development of independence, self-discipline, persistence, and love of learning. General IQ gains are not necessarily expected, at least not in one year, but children should make gains with respect to independence, persistence, and task-oriented attitudes. Increased self-control and respect for materials and the rights of others may also be expected.

Montessori anticipated current findings in characterizing the environment of the disadvantaged child as lacking in order and structure in comparison with that of his middle-class peers. She related the disorganization in the environment of these children to their handicaps in conceptualization and learning skills. In contrast, however, to much modern thinking with regard to the nature of the child, Montessori stressed the innate pride in achievement, the curiosity and high motivation to learn which characterize the preschool child, whether disadvantaged or not. She stressed the individuality of each child with his peculiar combination of capacities, fund of information and ways of learning. She believed the preschool child to be capable of intense and lengthy concentration.

2 Curriculum Content and Organization

The curriculum is organized into three large categories: exercises for daily living, sensorial materials, and academic materials. Normally exercises for daily living would form the beginning of the Montessori program and would be essentially the curriculum offered to three-year-olds, but four-year-olds without previous school begin here too, and the extent to which they progress to sensorial and academic materials is the function of each child's capacity to work through the program.

With respect to sequencing of tasks, Montessori much more resembles Bereiter-Engelmann than it does the Traditional program. It might be said that both Montessori and Bereiter-Engelmann styles involve programmed activity, but with one important difference. In Montessori the program is not imposed on the child, but is simply inherent in the nature of the materials and the ways in which they can be used. Sequencing is extremely important in Montessori and even the exercises for daily living are carefully programmed in small steps.

The Montessori classroom should contain children age three, four, and five. One reason for the mixture of ages in the preschool class is that younger children are expected to imitate older children in their behavior.

With the exception of short periods of whole-group activity, there is no formal grouping in the Montessori classroom. In fact informal grouping should occur infrequently since children are expected to work individually, and few of the materials are constructed for use by more than one person at a time.

In Montessori the child himself decides what he will study. The key term is self-education. Montessori teachers should be even less obtrusive than the teacher in the Traditional classroom. The teacher, however, is not passive. She should keep careful records on all individual children since it is her task to introduce new materials at the appropriate time. The appropriateness of the time is determined by the child's progress up to that point. This is what Hunt calls "the problem of the match" (1961).

Classroom atmosphere is extremely quiet and orderly. There is little emphasis on the emotional relationship between the teacher and the child. The teacher is not seen as a mother substitute but rather as an aide and a resource to the child in the process of self-education. Her manner should be friendly but somewhat detached as the child is supposed to be developing independence and the ability to direct his own activity. In this respect Montessori more resembles Bereiter-Engelmann than Traditional. The teacher does not interrupt children even to help them unless requested, or unless they are obviously in difficulties from which they cannot extricate themselves. There is heavy emphasis on orderly placement, proper use and care of materials, non-interference with others, and self-discipline.

3. Methods and Techniques

As is true with the Traditional program, Montessori emphasizes the difficulty that language presents in the learning of the very young child. These teachers are instructed not to talk any more than is necessary. There is no provision in the Montessori program for the remediation of linguistic deficits, though there is no proscription against the use of specific language materials where they seem to be called for. In general, however, this should not increase the amount of linguistic interaction occurring otherwise in the classroom, which is minimal.

On the question of reinforcement the Montessori program is quite different from the other three. The basic attitude derives from a belief in the spontaneous interest and joy which preschoolers take in learning, provided they are given an opportunity to attempt tasks which are suitable for their capacities.

Montessori insisted that children are annoyed by superfluous extraneous reward for something which is its own reward, namely mastery of a task. If the task is too easy the child will be bored; if it is too difficult he will be frustrated, but if it is just right he will enjoy the challenge and take pleasure in success. Negative feedback is expressly forbidden. The child is never to be told he is mistaken or wrong. He is simply to be reinstructed.

Sensory stimulation - the development of the senses - is considered vitally important in the intellectual development of the child. This focuses attention on the perceptual environment and on the materials to be presented to the child rather than on the child's response.

It is difficult to define the role which repetition or practice plays in the Montessori program. Although there is no explicit attempt to make sure that a child continues to repeat activities, the fact that the materials can be used in a variety of ways, some more advanced than others, does insure a certain amount of repetition. The child is free to initiate an activity or not and no pressure is put on him to do any particular thing. However, there is strong encouragement to follow a standard procedure once an activity is initiated and, thus, practice. It is probably fair to say that the emphasis is on sensory stimulation rather than on development of response habits since the habits emphasized are more procedural than "correct response habits" in the Hullian sense.

Manipulation of materials is a very important technique in Montessori--both from the standpoint of providing sensory stimulation and as a primary method of learning specific concepts. For example, children feel shapes, trace sandpaper letters, place cylinders in holes, identify objects by touch while blindfolded, etc. In addition, much of the academic curriculum is centered in the didactic nature of the special materials. Items such as knobbed cylinders are constructed to be self-correcting in that the child can observe directly through his own senses whether or not the task has been completed accurately.

Imitation (matched-dependent behavior) is very much a part of the Montessori program. Teachers show children in great detail each step in advance and expect the children to imitate exactly what they are doing. It is a major technique in skill learning and is used in all activities from washing dishes to manipulating counting beads.

Competition has no place in the Montessori program, nor is a child ever compared with another child. Total emphasis on uniqueness, individuality, and a respect for the child's own interests is vital.

Traditional Program

1. Philosophy and Goals

The goals of the official Head Start philosophy are very broad. They include cognitive, motivational, social and physical development. Children are expected to make gains in ability to think, understanding of the world, improvement in language skills--both receptive and expressive, curiosity about themselves and the world, more positive attitudes toward following instructions and imitating the teacher, greater need for achievement and a sense of pride in their own accomplishments, self-confidence, self-discipline, and the ability to interact successfully with one's peers and adults. This program is not preparatory, but focuses rather on development in many areas at each child's natural pace. This prevents setting any standards of achievement or specific goals for all children. The temporal focus is long-term. One source writer cautions against trying to make preschool a watered-down first grade. This is referred to as the "dribble-down disease".

The disadvantaged child is seen as not essentially different from any preschooler except that some characteristics are exaggerated. These children are said to be more in need of affection, less confident in themselves, lacking in experience with the environment, lacking curiosity. They are egocentric but at the same time insecure. They are eager to please, easily manipulated and likely to be damaged emotionally by acceleration, pressure, or over-control. They are very practical and concrete and their ability to utilize language is minimal. Disadvantaged children are often in poor physical condition and their language development is inadequate. They have not had stimulating experiences or materials necessary for adequate development.

2. Curriculum Content and Organization

The curriculum in the Traditional program is distinguished not by any particular content but rather by its flexibility. Broadly speaking, there is considerable similarity in the content of all preschool programs regardless of method. Consensus arises from the fact that there are many basic things that four-year-olds do not know--the names of common objects, basic concepts such as time, foods, etc., words used in making sensory discriminations in various modalities and many other things form a part of the curriculum for all preschool children. In the Traditional program the content may consist of anything which is of interest to the children at a particular time.

Emphasis in this program is on the relatedness of information in all areas. No sequencing of activities is necessary for no task is so foundational that all others depend on it. It is not considered crucial that certain facts or skills or concepts be mastered first.

In the Traditional classroom, the children are not grouped arbitrarily with the exception that there is a period during the day when they are all brought together for some whole-group activity such as singing or a story. Even at this time, however, no child is forced to participate. The so-called "free play" period, sometimes called "work-play" period occupies the largest single portion of the school day. This is a time during which the children are allowed to engage in whatever activities they choose and appears to be a time when they are simply playing. But the philosophy of the Traditional preschool is that at this age children do learn best through play. They are assumed to be planning, investigating, organizing ideas and developing skills. It is not clear whether the kind of play in which the child engages has any necessary relationship to what he learns. The material and physical arrangements in the Traditional preschool are intended to encourage various kinds of activities: pretend games in the housekeeping corner, physical exercises on balance boards and jungle jims, manipulation of materials such as puzzles which develop eye-hand coordination and sensory discrimination, and curiosity at the science table.

The teacher should be unobtrusive. Rather than trying explicitly to teach something she should provide stimulus situations and watch for opportunities to expand the children's horizons. Primary importance is placed on motivation and the enthusiasm which an event generates in children. Great stress is placed on the concreteness of the preschool child and the fact that whatever he learns must be related personally to him in some way.

The atmosphere in a Traditional classroom should be one of happy freedom within limits. Children should not be required to sit still for long periods nor should they be regimented. The class should be conducted at a leisurely pace. Particular stress is placed on the emotional needs of the children and the necessity for the teacher to be warm, patient, affectionate, tolerant and non-demanding. In this program the emphasis is on understanding and reaching the child, not on manipulating his behavior.

3. Methods and Techniques

There is a peculiar ambiguity in the Traditional program regarding the use of language. On the one hand the linguistic deficiencies of the disadvantaged are emphasized and stress is placed on the necessity to help children progress toward a more efficient use of language in both expression and listening. On the other hand the limited capacity of preschool children to use language in learning is emphasized. Teachers are cautioned not to talk too much nor to insist that the child speak more loudly or more distinctly because this may destroy his self-confidence. It would probably be accurate to say that the child's listening skills should be enhanced through the whole-group activities such as listening to records or listening to the teacher tell a story, and that his expressive skills should

improve as a result of conversation with the adults in the classroom and perhaps with the other children during the course of his play activities.

With respect to reinforcement, teachers are advised to give much praise but not to point out errors. They are specifically cautioned against the emphasis on right and wrong. There is no provision for material rewards nor any emphasis on a contingency between the child's performance and whether or not he receives reinforcement, though obviously in behavior management there is some contingency since the teacher does not praise undesirable behavior.

Sensory stimulation is strongly emphasized in the Traditional program. The child should have an opportunity to see, hear, taste, and manipulate many different things. Appropriate techniques include concrete items, field trips, visitors, pictures, fragrant items, movies and TV. An object-rich environment is provided.

The role of practice in learning is virtually unmentioned in any source material on the Traditional program. Repetition of sensory items is mentioned briefly by Hymes who says that young children learn slowly and need things repeated numerous times. No emphasis is placed on repetition of particular skills or response habits.

It is difficult to specify the role which manipulation of materials plays in this program. The environment and wealth of materials provided certainly promote physical interaction and handling. On the other hand, there is no explicit attempt to insure that children use materials nor is there any particular way in which a given item must be manipulated.

The role of imitation is not matched-dependent behavior as described in Miller and Dollard but is rather closer to the Freudian notion of identification, that is, "the endeavor to mold a person's own ego after the fashion of one that has been taken as a model". According to Jerome Kagan, two major goal states are involved in identification behavior. One is mastery of the environment and the other is love and affection. This notion of identification seems to be very much a part of the Traditional program. The teacher should become a much loved model of appropriate behavior. This meaning of identification appears to be more relevant to the development of values and attitudes than to the learning of skills. Children are not compared with one another and the use of competition is specifically proscribed.

C. TREATMENT IMPLEMENTATION

Teacher Training

From a pool of 22 of the previous year's Head Start teachers who indicated a willingness to attend special workshops during the summer in order to teach in the experimental programs, a cadre of 12 were selected. Criteria for selection were: (1) interest in program, (2) previous academic training and other indications of potential for new learning, (3) balance of such criteria in the three programs, and (4) likelihood of completion of the training program.

The four teachers who were to use the Traditional curriculum were sent to the regular 8-week Head Start training program for the region which is at the University of North Carolina; four were sent to an 8-week workshop at George Peabody College, where the DARCEE (Early Intervention) Program was developed; and four were sent to the University of Illinois for a 4-week workshop in the Bereiter-Engelmann method. Despite intensive efforts, it was impossible to recruit more than two Montessori teachers. These two had completed their Montessori training in the 8-week summer workshop at Fairleigh Dickinson University. Teaching the experimental classes constituted their required year of internship.

The assignment of aides was dictated by the placement of classes in schools, because Head Start guidelines require that aides be residents of the neighborhood in which Head Start classes are located. Consequently, the placement of a particular class determined who would be the aide.

Two-day workshops were held for aides in all the experimental programs at the University of Louisville just prior to the opening of school.

Bi-weekly meetings were held separately with teachers from each experimental program throughout the school year. These meetings were utilized more for the purpose of giving the teachers in each program an opportunity to communicate with each other than for the purpose of in-service training. Various problems having to do with general situations rather than specific program implementations were discussed. In addition to these meetings, arrangements were made for consultants from the various program styles to meet with their teachers twice during the year for two days each time. These visits from consultants were structured by them in accordance with their perception of the teachers' needs at the time. For the most part they consisted of observations and workshops as well as individual consultation.

Classroom Operation

All of the classes began in the first week in September and continued throughout the school year. The class day was 6½ hours long and all children were given a morning snack and lunch. They rested for a period of approximately one hour in the afternoon. In

addition to the basic furniture and materials used in all Head Start classrooms, classes were equipped with all special materials and supplies suggested by consultants as being of value in the implementation of their particular programs. For the Bereiter-Engelmann program, the Distar Reading Program was purchased and classes were provided with materials for the language and arithmetic programs. Total cost of equipment and materials was $1,487.00--a cost of $371 per classroom. Special equipment for the DARCEE program consisted primarily of Ginn Language kits, the basic five--beads, parquetry blocks, puzzles, peg boards, and counting cubes--and numerous books. Extra supplies were necessary for use of the two Home Visitors attached to these classes. These materials were left in the homes of the children for short periods. Total cost of equipping this program was $1,356--a cost of $339 per class. Montessori materials were ordered from the Netherlands. The total cost of equipping the two Montessori classrooms was $1,236--a cost of $618 per class. Books, puzzles, and other materials were purchased for the Traditional classrooms. Total cost was $1,138--a per class cost of $284.

Throughout the year teachers provided the research staff with attendance records, records of parent contacts, and of visitors to classes.

Parent Involvement

The usual parent participation emphasis in Head Start provided for regular meetings at each school, and teachers in all programs were urged to involve the parents as much as possible in the goals and activities of their particular program. Within the DARCEE program, the two Home Visitors represented a special effort to extend the curriculum goals into the home. These two teachers, selecting approximately half of the parents in each of the four classes, visited in the homes once a week, taking with them the materials currently being used by the teacher and leaving them with the mothers.

In order to assess the reactions of parents to the experimental program at the end of the year, a Parent's Evaluation Form was devised and sent to the parents of all children in each of the four experimental programs.

D. ASSESSING TREATMENTS

In attempting to compare educational programs one is immediately confronted with the fact that the independent variable is composed of a number of other complex variables. Verbal descriptions of the four programs compared in this study revealed many areas of agreement and disagreement among them. These programs apparently differ with respect to philosophy of education, goals of preschool, content and organization of curriculum, materials, methods and techniques, and in many other respects. However, all of these program aspects must be translated into teachers' classroom activity, where their effects, if any, will be manifest. Training programs, no matter how long or how thorough, cannot insure that teachers will implement what they have been taught. Verbal descriptions are therefore essential but inadequate. Labeling a classroom "Montessori" or "Traditional" may be only slightly more informative than labeling therapies Freudian or non-directive, unless the labels are supported by evidence of appropriate implementation.

Verification of Intact Treatments

One way of providing a definition of the "treatments" variable is to devise a method of answering the question, "Did the teachers really implement the programs in which they were trained?" In this form, the question is one of whether "X" is "really X", and like all questions of validity, leads to a regress culminating in consensual agreement regarding protocol statements. The obvious method of answering the question in the present case was by means of an evaluation of programs and teachers by those who developed each program, or who were involved in the training of individuals in the various methods. For this purpose, a "Consultant's Evaluation Form" was devised which called for magnitude ratings on a 0-10 scale. Consultants were asked to rate each class on teaching techniques, materials and a number of other aspects of programs. Ratings were made twice--once using an absolute criterion (in comparison with the ideal program) and again using a relative criterion (meaning considering the limitations imposed on implementation by distant location and the absence of resources available to programs under strict control in their original setting). The purpose of using these two criteria was to insure greater consistency on the part of the raters by making them aware of the distinction and giving them all a similar baseline. The conditions under which implementations were made precluded the possibility of their being prototypes of the original programs; the primary purpose was to determine whether the classes as a group were reasonable approximations of the original programs, and how classes ranked within programs. Ratings based on the "relative criterion", therefore, are of primary interest.

Analysis of Treatment Dimensions

A second method of defining treatments is to provide an operational definition of them in terms of their dimensions. That is, one may ask, "Along what dimensions should teacher behavior, child behavior, and classroom activity vary if these program are implemented?"

There are several reasons why this approach has greater utility than the documentation of program identity. For one thing, if intact programs do produce different effects, it is important to know which of the differences among them produced these effects. The *obvious differences between Program A and Program B* may be superficial and unrelated to their effectiveness. When programs do not have different effects it may be because different methods are equally effective or ineffective; but it may also be due to the operation of components common to both.

The most important reason for attempting to analyze treatment dimensions, however, lies in the fact that educational intervention programs are basically longitudinal experiments. They are attempts to modify development by the application of certain stimulus conditions over a longer period of time than is possible in a laboratory. But they are nonetheless experiments, and as such they have the potential for contributing substantially to our understanding of learning and development. This potential will not be fully realized unless experimenters can succeed in specifying and quantifying their independent variables. Treatment assessment procedures were therefore intended to serve a dual purpose: (1) to verify treatment implementation by determining the extent to which these particular classes contained the essential components of prototype programs; and (2) to provide a link between educational intervention and cognitive development by focusing on variables crucial to theoretical positions. Such a procedure involves selection of dimensions common to all programs and amenable to quantitative assessment, and systematic monitoring of classes to determine the amounts of these dimensions which occur in each program.

1. Selection of Dimensions

Research has identified many variables which should be important in learning or intellectual development. Comparison of program descriptions reveals that the programs can be roughly rank-ordered along a number of these dimensions. If teachers in these programs really do what their programs recommend, there should be differences in the amounts of language instruction, modeling, imitation, role-playing, reinforcement, manipulation of materials, sensorial stimulation, and many other variables.

This quantitative approach eliminates from consideration any characteristic which is uniquely present in one or more programs, but not in all four. The scope of this particular study also ruled out program characteristics which are not available to direct observation on a periodic, rather than continuous basis, for example, sequencing. The complexity of educational programs is emphasized by the fact that well over 100 variables were identified in the process of developing the monitoring procedures! Many of these are highly correlated, though not of necessity linked. Others occur so infrequently in preschool classes that quantitative assessment is impossible. The remainder constitute a bewildering array of techniques and combinations of techniques, most of which have probably never been formally

incorporated into any program.

2. Construction of Monitoring Procedures

Although a number of instruments have been devised to monitor teacher behavior or classroom activity, none of them seemed entirely adequate to serve the purposes of treatment assessment in this study. It was therefore necessary to develop new monitoring procedures.

It soon became obvious that a tally sheet which included all important dimensions of teacher behavior and classroom activity would exceed the capacity of raters to observe and record during a limited period of time. Although video-taping provides a permanent record which can be re-examined as often as desired, it also has a number of disadvantages. Technical problems make it difficult (unless professional quality apparatus and personnel are available) to obtain an adequate picture of an entire classroom at a given time. In addition, a single audio channel is insufficient to provide undistorted records of both teachers and children because of the frequency of simultaneous or overlapping vocalizations.

For these reasons the decision was made to construct two monitoring procedures. One procedure involved video-tape samples of ten-minute duration, focusing on the teacher for the entire period, and including all variables which could be assessed in this manner. The second was a time-sampling procedure for use in direct observation of classes. This tally sheet was constructed to obtain information which is difficult to monitor from tape--children's behaviors, grouping, and the nature of activities in the class. For verification and comparison it also included some of the important aspects of teacher behavior.

(a) Video-Tape Monitoring Procedure

The video-tape procedure was based ; Social Interaction Scale (1950); a number of categor added and a finer differentiation of types of feedbac made.

The three sample tally sheets (Figu , and 4) indicate the categories which were tallied. The and "Asking" distinction refers to whether the teache esenting something to the children or attempting to nething from them. Column dimensions represent the used, and row dimensions indicate the content or subs whatever was being presented or elicited. For inst cell marked "20" collects all instances in which th gave academic information verbally; the cell marked " ts all instances of teachers making a direct request for n. It should be noted that all possible acts which c r would be tallied with this procedure. Acts not o defined as techniques were coded under "peripheral

TEACHER'S ACTIVITY	VERBAL (V)	N.VERBAL (N-V)	EXEMPLARY (X)	MANIPULATORY (MANP)	MODELING (MOD)	INTERACTIVE (R-P)
HELP						
OPINION (OP)						
GENERALI-ZATION						
PROCEDURAL INFORMATION (P-I)						
ACADEMIC INFORMATION (A-I)	20					
CLARI-FICATION						
CONFIRMA-TION						
DISCONFIR-MATION						
STIMULA-TION						

FEEDBACK ACTIVITY	Knowledge of Results (K.O.R.)	CONTINGENT (CNTG)	NON-CONTINGENT
VERBAL & SIGN. (V)			
PHYSICAL			
MATERIAL			
ACTIVITY			

Fig. 2. Video Tally Sheet - Giving.

TEACHER'S ACTIVITIES	INDIRECT REQUEST (ID-R)	DIRECT REQUEST (D-R)	COMMAND
HELP			
OPINION			
GENERALIZATION			
PROCEDURAL INFORMATION (P-I)			
CONTENT INFORMATION (C-I)			
CLARIFICATION			
CONFIRMATION			
DISCONFIRMATION			
STIMULATION			
IMITATION (IM)			
ACADEMIC VERBAL PERFORMANCE (AVP)			
ACADEMIC NON-VERBAL PERFORMANCE (AN-VP)			
CONDUCT (OTHER)			

Fig. 3. Video Tally Sheet - Asking.

SETTING STANDARDS (SS)	
STATES REINFORCEMENT CONTINGENCY	
CITES PRINCIPLE	
CITES OTHER CHILD	
CITES TEACHER	
CHALLENGES	

PERIPHERAL ACTS	
OUT OF CONTACT ----(PA-NI)	
IN CONTACT BUT NOT INTERACTING ----	
CONVERSING (PA-CONV)	

Fig. 4. Video Tally Sheet - Miscellaneous

Most of the categories are self-explanatory, but some require elaboration. "KOR", "Contingent Reinforcement" and "Non-Contingent Reinforcement" are differentiated as follows: If feedback involved the right-wrong dimension, it was coded KOR. For example, "What is this sound? --- Right! It's *mmm*." If reinforcement was given for specific behavior, it was coded contingent reinforcement. For example, the child has finished a puzzle or completed a block design and the teacher says, "That's good. You put all the pieces in." Non-contingent reinforcement is simply praise which is general or which does not have anything to do with a standard. For example, "You were a good boy today" or "That's a very pretty dress". Exemplary ("X") as a technique is defined as involving the use of visual or auditory aids as a primary mode of instruction; manipulatory ("Manp.") involves the arrangement of objects such as tower-building, bead chains, etc.; modeling ("Model.") involves the use of the self or capacities of the self; interaction (role-playing) ("RP") was coded for any technique when it involved the teacher in an assumed role at the child's level, including dramatizations, pantomine, and table games. The distinction between content information ("CI") and procedural information ("PI") is that procedural conveys directions or instructions about how to do something whereas content infomation is the conveying of facts. Academic verbal performance ("AVP") is the practice of language per se.

Video-tapes were made of each class five times during the year. These tapes focused on the teacher, who was followed closely for a period of at least 10 minutes. Teachers wore the microphone around their necks and the transmitter around their waists. They were not connected in any way to the camera and were free to move around the room and follow their customary procedures. Teachers were not notified of the exact time at which tapes would be made, except that they expected it to occur during the week set aside for monitoring. They were instructed to continue their plans, regardless of whether a particular activity happened to be representative of their program. No attempt was made to maximize differences by having, for example, Bereiter-Engelmann teachers conduct patterned-drill during taping. The goal of the taping was to obtain a random sample of teacher behavior throughout the year.

(b) In-Class Tally Sheet

The in-class tally procedure (Figure 5) was a time-sampling method which assessed indices of the number of groups in classes, size of groups, shifts in group size, relative proportion of kinds of groups--whether doing different things (D/NF), the same thing (S), or engaged in a common enterprise (D/C), and total activity of all kinds tallied. This procedure also assessed seven categories of "teaching techniques" which were tallied for teachers, aides, volunteers, and children.

School:_____ Teacher:_____ Date:_____

Program:_____ Aide:_____ Observer:_____

Start:_____ Stop:_____

GROUP	ACTIVITY	MEDIA	GOAL	TEACHING TECHNIQUE							
A: Tally →				A	MANP	I-V	X	MO	RP	PG	CON
Size:				T							
				A							
Type: D/NF D/C S				V							
				C							
				Tot.							
B: Tally →				B	MANP	I-V	X	MO	RP	PG	CON
Size:				T							
				A							
Type: D/NF D/C S				V							
				C							
				Tot.							
C: Tally →				C	MANP	I-V	X	MO	RP	PG	CON
Size:				T							
				A							
Type: D/NF D/C S				V							
				C							
				Tot.							
D: Tally →				D	MANP	I-V	X	MO	RP	PG	CON
Size:				T							
				A							
Type: D/NF D/C S				V							
				C							
				Tot.							
E: Tally →				E	MANP	I-V	X	MO	RP	PG	CON
Size:				T							
				A							
Type: D/NF D/C S				V							
				C							
				Tot.							

Fig. 5. In-Class Tally Sheet.

Manipulatory ("MANP") and role-playing ("RP") include the same activities as in the video-tape procedure. Verbal instruction ("IV") collects instances of academic instruction given verbally or recitation by children. No distinctions were made on the in-class procedure as to whether verbal instruction was used to convey content or was used for language practice. Exemplary ("X") is broader than the exemplary category in the video-tape procedure. It collects all instances of showing or being shown for instructional purposes, including pictures, objects, designs, sounds, fragrant items, movies, TV, songs, records, music or persons; it includes rhythms demonstrated by the teacher, and writing on the chalkboard in connection with instruction. Much of the activity coded "Modeling" on the video-tape was coded under the "X" category on the in-class tally sheet. Motor activity ("MO") was used only when some large muscle activity was going on, such as riding a tricycle, walking a balance board, playing ball or any active games, marching, doing exercises, playing outdoors, playing on the jungle jim, etc. Motor was not coded simply because children or teachers were moving around, but only when this was the principle technique occurring at a given time. Physical guidance ("PG") was intended to be coded whenever any adult guided the child manually through an activity--e.g., taking the child's hand and guiding him in drawing a line or taking his foot and moving it on the pedals of a tricycle. Since it involves the child being passively manipulated, it was not to be coded for children. Conversation ("CON") between children was coded after "child", except for fragmentary remarks incidental to activity which were not coded.

Two scores are available on Teaching Techniques. (1) The amount of each technique relative to the total number of acts tallied. This is an index of the frequency of use of a given technique relative to other techniques used (designated "Cell/Row"). (2) The absolute amount of each technique as a proportion of the number of times tallying was done (number of 15-second periods). This is an index of how often a given technique was used regardless of the frequency of other techniques (designated "Cell/Tally").

For example, if a teacher did very little verbal instruction in comparison with other teachers, her cell/tally percentage would be low. But if she used verbal instruction a lot more than she used any other technique, her cell/row percentage would be high.

In contrast, a teacher using large amounts of verbal instruction in comparison with other teachers, but also many other techniques about as often as verbal instruction, would be high on cell/tally but low on cell/row.

After preliminary agreement was obtained on categories and procedure by classroom tallying and subsequent discussion, three rater-reliability studies were done, two using simultaneous monitoring of tapes, and one using simultaneous monitoring in a classroom. Chi square was non-significant for all sessions, the last one being only .81, suggesting that monitors were consistently recording events in the same categories. Tallying was done by five monitors, who rotated among programs and classes over five sessions of two hours each. The monitor entered the class and identified all groups. A group was defined as one child alone or a number of children in close physical proximity. The number and type of groups were recorded. The monitor then observed each group for a period of 15 seconds, tallying every instance of all techniques, but no more than once for each, during the 15-second period. When all groups had been observed, the monitor returned to the first group, noted changes in size or activity, and tallied for another 15 seconds, etc.

E. ASSESSING TEACHERS

Attitudes, personality characteristics, and teacher intelligence may have direct effects on children's performance but in addition they may interact with the effects of training programs. Although it would be impossible to unravel all those complexities in one study, an attempt was made to obtain some information about teachers by means of questionnaires and tests.

Personality

Personality was assessed by means of the 16 Personality Factor Questionnaire.[1] This inventory assesses 16 primary bi-polar factors and in addition four secondary factors as follows: low anxiety vs. high anxiety, introversion vs. extraversion, tenderminded emotionality vs. alert poise, and subduedness vs. independence.

Intelligence

The Peabody Picture Vocabulary Test was used to assess teacher intelligence. This test lacks the threatening qualities of an IQ test such as the Binet. It provides a measure of vocabulary which is the best single estimate of IQ.

Attitudes towards Teaching

To assess teachers' attitudes toward teaching and toward children, a questionnaire developed by Neill (1967) was given to all teachers. This questionnaire was designed to assess attitudes in five areas: toward teaching, toward knowledge, toward peers, toward self and toward pupils.

Agreement with Program Philosophy

In order to obtain some information on the degree to which the training program had succeeded in orienting teachers towards particular programs, a "Statements Test" was devised in the following way: A number of statements regarding various aspects of these preschool programs were extracted from the publications and source materials in each program. These statements were then typed on 3x5 cards and presented to consultants who were asked to rate them on a 5-point scale from strongly agree to strongly disagree. This Statements Test was given to all teachers in December 1968, and January 1969.

---

[1] The Institute for Personality and Ability Testing
1602 Coronado Drive, Champaign, Illinois 61820

F. ASSESSING TREATMENT EFFECTS

Selection of Instruments

The primary criterion for selection of the main battery to evaluate the effects of treatments on children's performance was the necessity to tap a wide range of dimensions in which change might be expected to occur. This decision was dictated partly by the varied goals of the programs and partly by the paucity of knowledge in the area of compensatory education as to what changes might be most lasting or effective at later periods.

Four major dimensions of development were chosen: cognitive, motivational, perceptual and social. A fifth very important dimension was eliminated, namely personality. The lack of suitable instruments to assess personality variables in four-year-olds and the necessity to limit testing time for children of this age made impracticable the attempt to measure such variables as ego strength and anxiety in addition to the areas more obviously related to academic progress.

Five additional tests were administered at the end of the year to a sample of six children from each class--primarily to assess specific skill learning. These tests are described separately from the main battery under "Additional Tests".

1. Cognitive Variables

Stanford-Binet, Revised, 1967 ("S-B")[1]

The decision to use the Stanford-Binet as a measure of intellectual functioning was supported by a number of considerations, among them the fact that the Binet is, to date, the best predictor of school achievement, and is probably the best single test of global IQ. The wide use of the Binet in studies assessing the value of various programs for preschool children was an additional argument for its inclusion. Program developers have themselves used it to assess the Bereiter-Engelmann and DARCEE programs.

The Preschool Inventory ("PSI")[2]

Although it has a high correlation with the Stanford-Binet, the Preschool Inventory was selected because of the four sub-tests in the standardization version, representing factors for which we had no other instruments of assessment. Although these factors, Personal-Social-Responsiveness, Associative Vocabulary, Concept Activation-Numerical and Concept Activation-Sensory, do not appear on the 1968 revised version of this instrument which was used, it was hoped that results would be analyzable in terms of these sub-test factors. The 1968 Experimental Edition of this

[1] Houghton Mifflin Company, 666 Miami Circle, N.E., Atlanta, Ga.

[2] Educational Testing Services, Princeton, New Jersey 08540

instrument contains 64 items, a considerable reduction from the original 85.

Quick Picture Vocabulary Test ("Q")[1]

The selection of this instrument represents an attempt to assess intellectual functioning by means of a test which does not penalize the disadvantaged child. The authors provide evidence that the Quick is "culture fair" in this sense (1962). The Quick is very easy to administer and also very short. The child is given a card on which there are four pictures. As the examiner speaks a word, the child's task is to select from the four pictures the one which best represents that word.

2. Motivational and Social Variables

The Curiosity Box ("C-V" and "C-A")[2]

This test is precisely what the name implies, a box containing a variety of items inside and outside which the child can manipulate or look at. Two scores are obtained--a score for verbalization regarding the box and a score for actual exploration (activity).

The Replacement Puzzle ("R-P" and "R-R")[2]

This instrument was designed as a test of task-persistence and distractibility. The child is provided with a board on which there are a number of non-removable shapes and four shapes which can be lifted out. These four can be replaced in only one way so that they will lie flat. The time limit of three minutes makes this a very difficult task for most four-year-olds, and few of them are able to solve it within this interval. A distractor is provided at the end of two minutes and the child's score is based on the total length of time during which he is oriented towards solving the task both before and after distraction.

The Dog and Bone Test ("D-B")[2]

According to Banta (1968), this is a test of "initiative". The material consists of a small board on which are four wooden houses, one at each corner, a small dog at one end, and a bone at the other. The task is to devise a variety of paths over which the dog can travel in order to reach his bone. The score is based on the number and quality of different paths which the child is able to produce. This might also be considered a test of creativity, inventive thinking or some other aspect of cognitive style.

---

[1]Psychological Test Specialists, Box 1441, Missoula, Montana 59801

[2]Cincinnati Autonomy Battery, Dr. Thomas Banta, University of Cincinnati

Face Sheet of the Binet ("F-AC", "F-CN", and "F-AL")

The results of factor analysis by Hess, et. al., (1966) indicated that the Face Sheet can be used to assess achievement motivation, confidence in ability, and activity level.

Behavior Inventory ("B-AG", "B-VP", "B-TM", "B-IN", and "B-AC")

This is a rating scale completed by teachers. The version of the Behavior Inventory used is that recommended by Hess, et. al, (1966) consisting of 20 items which can be summed to form sub-test scores loading on five factors: Aggression, Verbal-Social Participation, Timidity, Independence, and Achievement Motivation. The original 4-point scale was used since these authors found the correlation between the 4-point and the expanded 7-point scale to be very high.

3. Perceptual Variables

Embedded Figures Test ("EBF")[1]

This is the fourth test from the CAB. The task is to locate a cone embedded in various line drawings, some geometric and some realistic. The child covers the cone with a duplicate cut-out. The author believes that it measures "field-independence". It involves a complex visual perceptual skill.

Wepman Auditory Discrimination Test[2]

The Wepman assesses differentiation on initial and final consonants and middle vowels. Although it has been standardized only down through the age of five, it was selected because at the time it seemed to be the only standardized test of auditory discrimination available which might be successfully used with four-year-olds.

4. Additional Tests (Sample Only)

Parallel Sentence Production ("PSP")[3]

This test requires the child to produce a complete sentence about a drawing which is on the same page. For example, the tester says, pointing to a drawing, "This small boy is riding a small bike". Then the tester points to the other picture which is a picture of a larger boy riding a larger bike and says, "Tell me about this picture". The child is given credit if he says, "This large boy is riding a large bike" or "This big boy is riding a big bike".

---

[1] Cincinnati Autonomy Battery

[2] Language Research Associates, 300 N. State St., Chicago, Illinois 60610

[3] UCLA Preschool Research Projects, Dr. Carolyn Stern, Director 1019 Gayley Ave., Los Angeles, California 90024

Expressive Vocabulary Inventory (

This test is similar to a Picture ry Test, but requires the child to produce a variet ts of speech, including prepositions, verbs, adjectiv.., etc., and also requires identification of portions of stimuli, and of collective nouns describing a number of objects. It also calls for comparisons. Examples are: What's the boy doing? (Swimming). This is a whole apple, what part of the apple is this? (Half). See this ball? It is smaller. What about this ball? (Larger or bigger). What are all of these called? (Animals).

Basic Concept Inventory ("BCI")[2]

This is a test requiring picture selection. It involves listening vocabulary, particular attention to words which change the meaning of sentences and also reasoning. For example, on one card the child must find the picture which is correct for the statement, "She is between a boy and a girl". On another card he must differentiate among pictures which correspond to these sentences, "The man is going to chop down the tree", "The man chopped down the tree", "The man is chopping down the tree". Other items test the child's knowledge of language structure with nonsense words. For example, "Fends cannot crump. Can fends crump? What can't fends do?"

Arithmetic Test ("ARIT")

Portions of arithmetic tests devised for use with children in Bereiter-Engelmann classes were combined and used as a test of simple counting and addition.

Peabody Picture Vocabulary Test ("PPVT")[3]

This test was given primarily in order to determine whether it and the Quick Test give similar results on four-year-olds. The Peabody has been used more often than the Quick in studies of preschool programs.

Recruitment and Training of Testers

Stanford-Binet testers consisted of advanced graduate students or professional psychologists who were experienced in the administration of the test to young children. Criterion was completion of at least one supervised practicum. In fact, however, all testers had completed at least two practicums in testing and many of them had previously tested large numbers of Head Start students. The Binet testers were given a one-hour orientation to acquaint them with research design

[1]UCLA Preschool Research Projects

[2]Follet Educational Corp., 1010 West Washington Blvd., Chicago, Illinois 60607

[3]American Guidance Service, Inc., Publishers Bldg., Circle Pines, Minn.

and procedures and orient them toward standard instructional procedures in testing. The remainder of the tests in the main battery--eight in all--were divided into two groups in such a way that total testing time for each group of tests was approximately 20 to 40 minutes. This arrangement also had the effect of balancing the various kinds of tests within both groups. Group A consisted of the Curiosity Box, Embedded Figures, Wepman Auditory Discrimination Test, and the Quick Picture Vocabulary Test. Group B consisted of the Replacement Puzzle, Dog and Bone, Preschool Inventory, and the Face Sheet of the Binet. Within each group the test considered to be the most interesting to the children was given first. In Group A this was the Curiosity Box; in Group B, the Dog and Bone.

A number of graduate students in Psychology and Education were recruited and trained to give these tests--half on Group A and half on Group B tests. The training program consisted of 12 hours. Two seminars were held on the tests, the administrative procedures, scoring, methods of handling four-year-olds, general testing problems, and the purpose of the research. All testers were then required to give all their tests to at least three four-year-olds. Arrangements for practice testing were made with a private preschool. Following the practice testing, a final seminar was held at which various questions and problems were discussed.

Two of the testers were selected for training on the additional tests to be given at the end of the year.

Procedural Controls

Three controls with respect to test administration were considered essential. These were: the order of test presentation to children, the interval between first and second tests, and the distribution of testers among programs.

1. Order of Test Administration

All subjects were given both Group A and B tests before they were given the Stanford-Binet. Thus by the time the children were given the Binet, they had taken seven other tests given by two different testers and were fairly sophisticated regarding testing procedure. This arrangement in addition to the 6-week postponement of testing to allow for school adjustment was an additional effort to minimize the "testability" factor for the Binet.

Although it was not possible to achieve strict counterbalancing of the order of presentation so far as Groups A and B tests were concerned, the order was scrambled with respect to programs. Within Groups A and B the order of testing remained the same for all subjects.

2. Interval Between Tests

Although the school year continued for nine months, the necessity to allow a period of adjustment to school at the

beginning of the year and the total testing time involved on both ends shortened this interval considerably. The goal was a first-second test interval of approximately six months. In order to maintain this same interval of time between testings for all subjects, it was necessary to give both the fall and the spring tests to classes in the same order insofar as this could be done. In general, the same order was followed in both testing sessions--among and within classes.

3. Distribution of Testers Across Programs

Ten individuals did Binet testing and the schedule was arranged so that there was no systematic bias with regard to programs. For the bulk of the testing it was possible to rotate testers among programs. Testers were uninformed regarding the nature of particular classes.

Groups A and B testers entered classes and tested all available children then moved on to another class, sometimes within the same school, sometimes in another school. The testing schedule was arranged so that testers alternated programs as they finished classes. One group of testers followed on the heels of the other, completing classes as they were vacated. These testers were also uninformed regarding the nature of the programs in individual classes.

Control children were tested last--both in the fall and again in the spring.

Rater Comparisons

1. Face Sheet

Hess et. al., (1966) report a sizeable correlation between children's IQ on the Binet and the Face Sheet ratings made on the children by these testers. This suggests that experienced testers may be influenced in their Face Sheet ratings by estimates of the child's IQ. In order to provide some additional information on this relationship and to check on the reliability of the ratings, the Binet Face Sheet was completed by the testers who administered the Preschool Inventory, and also by Binet testers.

2. Behavior Inventory

The Behavior Inventory was completed by the aides as well as the teachers in order to provide a comparison of independent ratings of subjects by different individuals in the classroom.

IV. PREDICTIONS

Two categories of predictions were made: (A) Predictions about the dimensions of classroom activity as a function of programs, and (B) Predictions of treatment effects.

A. *Treatment Dimensions*

It was predicted that language as an instructional method would occur most often in the Bereiter-Engelmann and DARCEE classes and least often in Montessori. It was also expected that teachers in the Bereiter-Engelmann program would elicit more practice in academic verbal performance than teachers in other programs. Conversation, on the other hand, was expected to occur most frequently in DARCEE and Traditional classes and least often in the other two.

It was predicted that the least manipulation of materials by teachers would occur in Bereiter-Engelmann, most in Montessori, and that manipulation by children would be high in Montessori, Traditional and DARCEE and low in Bereiter-Engelmann.

Motor activity and role-playing were expected to be high in the Traditional program for children. Role-playing was expected to be low in Montessori and Bereiter-Engelmann.

With regard to reinforcement, it was predicted that more of all kinds would occur in Bereiter-Engelmann and DARCEE classes than in the other two programs and, specifically, more knowledge-of-results in Bereiter-Engelmann and DARCEE. More negative KOR was expected in Bereiter-Engelmann than in the other three programs.

It was predicted that three groups or fewer would be found more often in Bereiter-Engelmann and DARCEE programs and that the number of groups in classes should be highest for Montessori and Traditional. Groups doing the same thing (S) were expected to be frequent in Bereiter-Engelmann and DARCEE while groups doing different things, but with a common group focus (D/C), should be found more often in Traditional. Groups doing different things with no group focus (D/NF) were predicted for Montessori. It was also expected that D/C would increase for Traditional toward the end of the year as children learned to work more cooperatively with each other. Relatively stable grouping in Bereiter-Engelmann, DARCEE and Montessori was predicted, with more frequent shifts within groups in Traditional. Groups in Montessori were expected to become more stable toward the end of the year as children learned to work for longer periods at individual projects.

B. Treatment Effects

Increases in IQ, achievement motivation, academic achievement and auditory discrimination were predicted for Bereiter-Englmann children; increases in IQ, persistence, resistance to distraction, achievement motivation, independence, auditory discrimination, and academic achievement were predicted for DARCEE. Montessori children were expected to score high in curiosity, persistence, resistance to distraction, initiative and independence. Traditional children were expected to be high in initiative, curiosity, self-confidence (lack of timidity), and verbal-social participation.

V. RESULTS

A. Design

The first results considered were those which provided evidence regarding the success of random assignment of subjects and experimental replications, the adequacy of the lower-class control group, and the success of procedural controls.

Table 1 shows that, according to demographic information obtained on all subjects, there were no significant differences among groups assigned to the four programs. It appears that programs were successfully balanced with respect to the variables generally thought to be important aspects of social class.

It is also clear that the lower-class controls were similar to the experimental population with the exception of two variables: the control group had a higher percentage of children who were living with both father and mother than was the case in the experimental group, and also a higher percentage of white children. Since approximately two-thirds of this control group were obtained from waiting lists for Head Start, while the remainder were children recruited otherwise in the community, comparisons were made between these two types of controls on all dependent variables. Only one significant t value was found out of 24 calculated. The conclusion was therefore reached that wait-listed controls did not differ from controls who had not been registered for Head Start classes.

With respect to procedural controls, Table 2 shows that the interval between fall and spring tests was the same within four or five days on all tests. Distribution of testers across programs is difficult to summarize in tabular form, but the salient results are that in only one of the 14 classes were children retested by the same tester, and in no program were fewer than two testers used. The number of different testers used in each program is also shown in Table 2.

Summary

In summary, random assignment of children to classes and balancing of the four programs by replication of the experiment in different areas appear to have been successfully accomplished and the lower-class control group obtained did not differ from the experimental group in any demographic characteristic which would be expected to favor the experimental children. This group was, in most respects, similar to the experimental population, whether controls were wait-listed for Head Start or not. It appears that the population of children in experimental Head Start classes was representative of four-year-olds in these poverty areas in the city of Louisville in most respects.

TABLE 1

Demographic Information on Program and Control Subjects

	Programs				
	Bereiter Engelmann	DARCEE	Montessori	Traditional	Controls
Means	N=64	N=64	N=33	N=52	N=34
Mean Age of Children (months)	51.40	51.76	52.68	51.07	52.61
Mean Income[a]	$2,943	$3,148	$2,886	$3,170	$3,680
Mean Age of Mother	27.98	28.77	28.00	28.05	29.73
Mean Age of Father	31.50	32.28	31.76	31.88	32.54
Medians					
Median education of Mother[a]	12	11	11	11	10
Median education of Father[a]	12	11	11	12	10
Median No. Siblings	2	3	3	2	3
Median No. in Home	5	6	5	6	6
Percentages					
% of Males[a]	42.1	50	54.6	39.6	52.9
% of Females[a]	57.9	50	45.4	60.4	47.1
% of Negro Children	96.9	87.5	100	88.7	73.5
% of White Children	3.1	12.5	0	11.3	26.5[b]
% Living with Mother and Father[a]	33.3	34.9	27.2	42.3	61.8[b]
% Living with Mother only[a]	58.7	57.2	66.7	57.7	35.3[b]
% Living with Father only	3.2	0	0	0	0
% Living with neither parent	4.8	7.9	6.1	0	2.9

TABLE 2

Fall-Spring Test Intervals and Number of Testers Used

Program	Group A Tests		Group B Tests		Stanford-Binet	
	Mo - Da	# Testers	Mo - Da	# Testers	Mo - Da	# Testers
Bereiter-Engelmann	6 - 1	(4)	6 - 6	(4)	6 - 4	(7)
DARCEE	6 - 2	(4)	6 - 7	(6)	6 - 5	(5)
Montessori	6 - 3	(3)	6 -10	(3)	6 - 5	(4)
Traditional	6 - 3	(4)	6 - 6	(5)	6 - 2	(5)
Controls	6 - 0	(5)	6 - 0	(2)	5 -27	(6)

B. Profiles

In order to provide an overview of the results on assessment of treatment dimensions and treatment effects, standard score profiles are provided. Figure 6 shows both absolute (Cell/Tally) and relative (Cell/Row) amounts of teaching techniques monitored in class for teachers and for children. Figure 7 shows the dimensions of activity for teachers monitored from video-tapes. Figure 8 presents treatment effects.

C. Treatment Characteristics

Verification of Intact Treatments

Results from the consultants' ratings of programs are shown in Table 3. Examination of this table indicates that all programs received ratings above the mid-point in respect to being demonstrations of their respective styles. Consultants from the Bereiter-Engelmann program were least pleased with implementation. There may be several reasons for this. For one thing, the training program which these teachers had was only four weeks long as compared with eight weeks in the other three programs. Secondly, there was some difficulty in obtaining on-site consultation on this program early in the year. Because of this a temporary confusion arose regarding the appropriate order of materials, and arithmetic programs, in particular, were not carried out in the proper way until late in the year. Finally, considering the highly specialized nature of the material and the close dependence of program implementation on these materials, this program may simply be the most difficult to implement without continual on-site supervision.

Consultants for the Traditional program were most pleased. It is noteworthy, however, that their rating of "Facilities" was quite low, despite balancing across programs. This may reflect a greater emphasis on aesthetic value in the Traditional program, or it may simply indicate that these consultants place more importance on facilities than do the consultants for the other programs.

DARCEE and Montessori ratings were second and third highest, respectively. Montessori classes were expected to receive low ratings, partly because there were only two of them, but primarily because Montessori classes composed entirely of four-year-olds violated a procedural standard--that is, the mixture of children of ages 3, 4, and 5.

Consultants' reports were also requested on the teachers and programs following the in-service training sessions. These were more useful as aids to in-service training than as evaluations. In general, however, they did correspond well with actual ratings given on the Consultant's Evaluation Form.

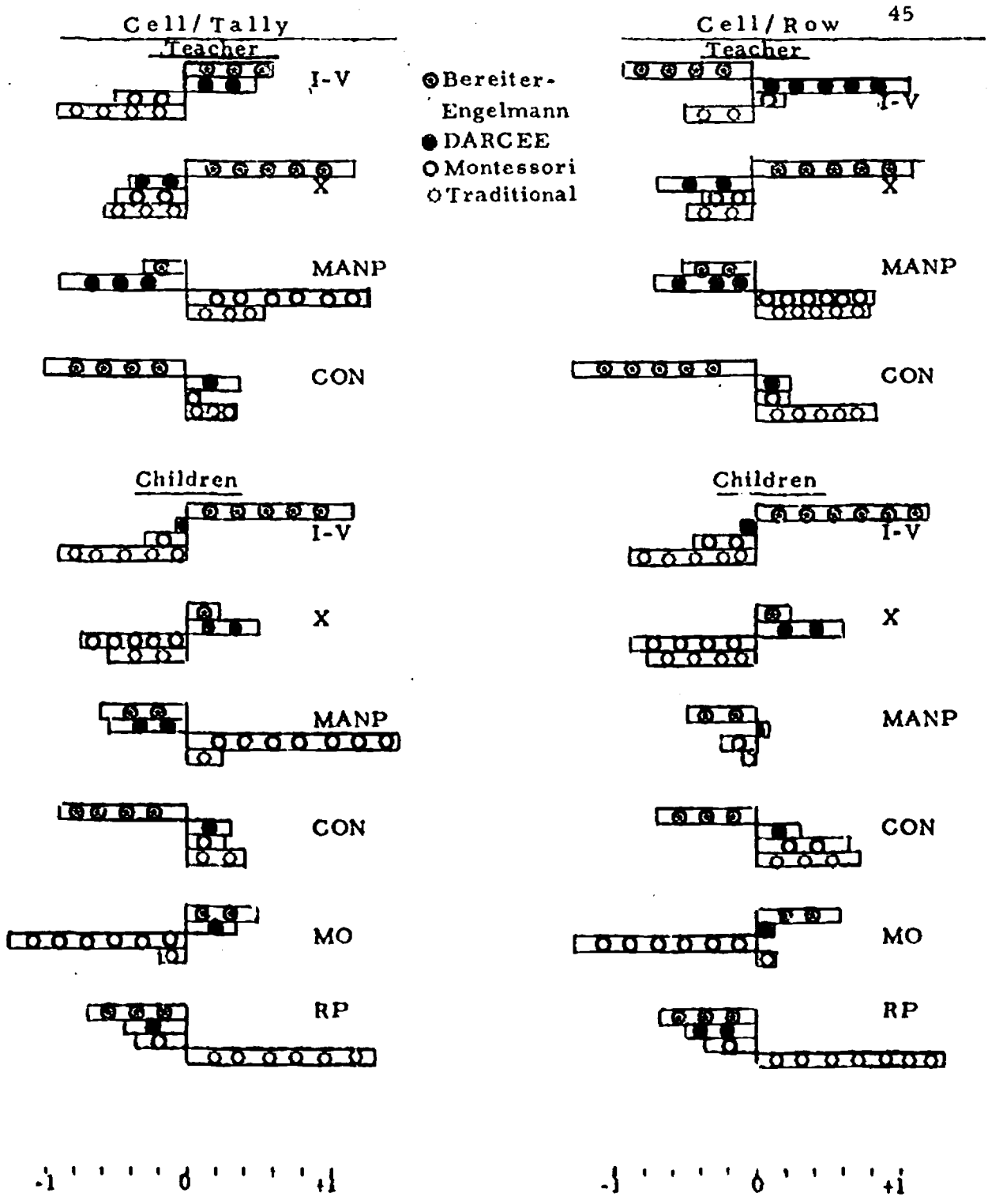

Fig. 6. Standard score (Z) profiles for in-class monitoring on teachers and children - Cell/Tally and Cell/Row measures.

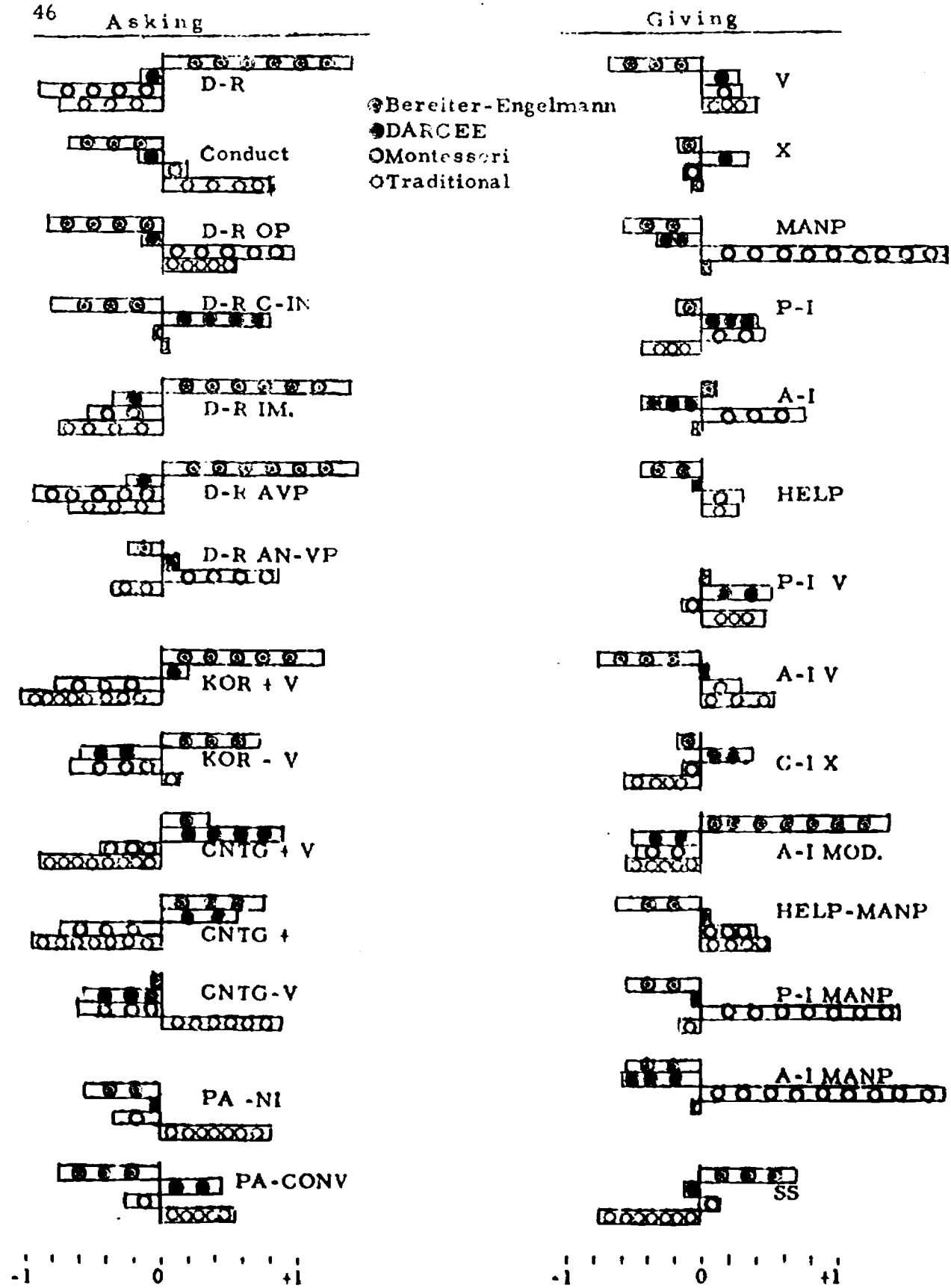

Fig. 7. Standard score (Z) profiles for Video-tape monitoring on teachers.

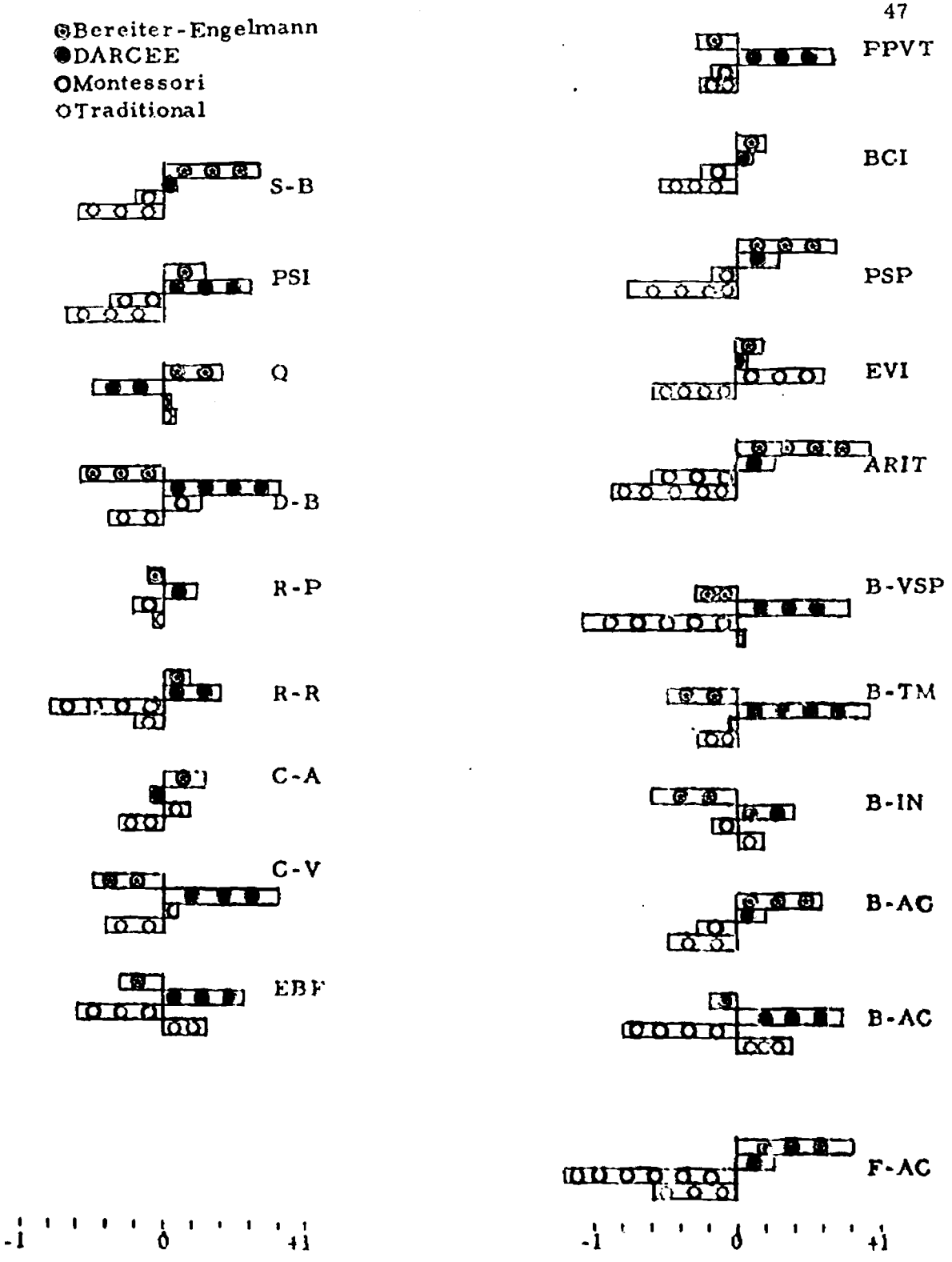

Fig. 8. Standard score (Z) profiles for treatment effects.

TABLE 3

Consultants' Rating of Programs

Program Aspect Evaluated	Programs			
	Bereiter-Engelmann	DARCEE	Montessori	Traditional
Teaching Techniques	5.87	7.28	6.25	9.17
Materials	7.50	7.00	7.50	10.00
Principles	5.62	7.86	6.50	9.00
Context	5.50	7.71	7.00	8.50
Selection of Activities	5.00	7.86	7.25	8.83
Content	8.50	6.71	6.50	9.17
Facilities	7.87	8.29	5.00	3.67
Progress of Children	5.86	7.25	6.25	-
Classroom events typical	7.00	7.43	6.50	8.67
Extent to which a demonstration	5.37	7.71	6.25	10.00
All Categories	6.33	7.51	6.50	8.56

Notes.- Means represent the relative criterion ratings. Means for classes are on a 0-10 scale ("Not at all" - "Best possible").

Analysis of Treatment Dimensions

Analysis of variance was used to determine differences among programs on the in-class monitoring variables. Three different analyses were completed in order to make all necessary comparisons: a 3x4 analysis comparing the Bereiter-Engelmann, DARCEE and Traditional programs in all four areas, a 2x4 analysis which included the Montessori program in two areas, and a one-way analysis of variance which eliminated the area factor but also compared all four programs. Tukey's multiple comparison procedure was used to compare differences between the means. For the 2x4 analysis, only significant differences between Montessori and other programs are reported. The arcsine transformation was used to reduce positive skewness on the Cell/Tally and Cell/Row proportions. A Kruskal-Wallis one-way analysis of variance was used to detect program differences in number of changes in group size and type of group. The chi square test was used for the number of groups. All differences are reported at the .05 level.

1. Teaching Techniques

(a) In-Class Monitoring

Table 4 presents the amounts of various teaching techniques and classroom activity for both teachers and children as monitored by the in-class procedure.

Manipulation. Cell/Tally percentages did not significantly differentiate teachers but Montessori teachers were highest. Cell/Row percentages showed that, relative to other techniques, Traditional teachers did significantly more manipulation of materials as a teaching technique than teachers in the Bereiter-Engelmann and DARCEE programs. This result was unexpected but probably reflects the low incidence of other techniques in the Traditional program.

Manipulation is more important as a technique in terms of its use by children than by teachers. Table 4 shows that Montessori children did significantly more manipulation of materials than children in Bereiter-Engelmann and DARCEE classes. Relative percentages were also higher for Montessori children than for Bereiter-Engelmann children, reflecting the fact that children in Montessori classes were usually manipulating materials whereas those in the Bereiter-Engelmann classes were usually reciting. Children in the Traditional program were second highest in absolute percentage--not significantly different from Montessori children in these two areas.

Verbal Instruction. Bereiter-Engelmann and DARCEE teachers used verbal instruction more often than teachers in the other two programs. Cell/tally percentages shown

TABLE 4

Means on In-Class Monitoring for Teachers' and Children's Behavior

Program	MANP.	I-V	X	MO	R-P	CON.
Teachers						
Bereiter-Engelmann						
Cell/Tally	.17	1.13	.99	Virtually zero.	Virtually zero.	.14
Cell/Row	.23	1.62	1.39			.19
DARCEE						
Cell/Tally	.12	1.08	.58			.28
Cell/Row	.20	1.95	.93			.44
Montessori						
Cell/Tally	.30	.89	.56			.25
Cell/Row	.49	1.79	1.01			.44
Traditional						
Cell/Tally	.24	.79	.52			.27
Cell/Row	.50	1.70	1.01			.54
Children						
Bereiter-Engelmann						
Cell/Tally	1.16	1.00	.38	.39	.18	.39
Cell/Row	1.34	1.21	.47	.51	.22	.43
DARCEE						
Cell/Tally	1.16	.56	.43	.35	.24	.48
Cell/Row	1.56	.69	.53	.47	.29	.69
Montessori						
Cell/Tally	1.81	.50	.25	.18	.27	.66
Cell/Row	2.00	.54	.26	.18	.29	.68
Traditional						
Cell/Tally	1.40	.30	.31	.36	.67	.66
Cell/Row	1.53	.33	.33	.40	.73	.70

Note.- Arcsine transformation.

in Table 4 for both of these programs are significantly higher than in Traditional, and Bereiter-Engelmann is significantly higher than Montessori in those two areas. Program order from most to least was: Bereiter-Engelmann, DARCEE, Montessori, Traditional. The relative percentages show that DARCEE teachers used a significantly greater amount of verbal instruction than Bereiter-Engelmann and Traditional.

For children the ordering of programs parallels the ordering for teachers; that is, from most to least: Bereiter-Engelmann, DARCEE, Montessori, Traditional. Relative percentages reflect the same order.

Exemplification. Bereiter-Engelmann teachers were significantly high in this category both absolutely and relatively. This result explains the rather low cell/row percentage for verbal instruction in the Bereiter-Engelmann program since in this program verbal instruction is almost always accompanied by showing the children something, usually a page in the teacher's manual. Thus, while the absolute amount of verbal instruction is high, it is not high relative to other techniques.

For children, absolute percentages on exemplification do not differentiate programs but DARCEE children are high. Cell/row percentages were significantly high for DARCEE children in the two areas containing Montessori classes.

Motor. There were no significant differences among programs in this category, though it is noteworthy that both relative and absolute percentages were highest in the Bereiter-Engelmann program.

Role-Playing. Role-playing by children was highest in Traditional classes and lowest in Bereiter-Engelmann. Traditional classes were significantly high in both absolute and relative percentages than each of the other three programs.

Conversation. The ordering of programs with respect to teachers' conversation with children was from most to least: DARCEE, Traditional, Montessori, Bereiter-Engelmann. The differences among programs did not reach significance at the .05 level for cell/tally percentages. Cell/row percentages, however, are similar and are statistically significant with DARCEE and Traditional being greater than Bereiter-Engelmann. Montessori teachers had as much conversation with the children relative to other techniques, however, as DARCEE teachers did.

For children, conversation includes both conversation with teachers and with other chilren. In absolute amount, the difference between greatest and least was not quite

significant, using the relatively conservative Tukey test; but there was more conversation in the Traditional program than in Bereiter-Engelmann as predicted. Surprisingly there was almost as much in the Montessori classes as in the Traditional classes. The amount in DARCEE classes was not especially high. Since teacher conversation with children was high in DARCEE it appears that most of the conversation which occurred in DARCEE classes was conversation between teachers and children, whereas somewhat more conversation among children occurred in Traditional classes.

(b) Video-Tape Monitoring

Table 5 presents means in various categories assessed by the video-tape monitoring procedure. Means comparisons have not been made, but analysis of variance indicates that the categories listed in Table 5 are significant sources of variance among programs. Because of very infrequent tallies in many cells, column and row categories were combined in various ways for the purpose of analyzing the frequencies. For example, almost all of the "asking" or elicitation coded fell under "Direct Request" rather than "Indirect Request" or "Command".

Contingent Positive Reinforcement. This category includes all positive reinforcement, verbal or material, which was given for specific behavior whether academic or otherwise, provided the reinforcement was contingent on meeting a standard. It does not include knowledge-of-results (KOR). Table 5 shows that Bereiter-Engelmann and DARCEE programs were high with a mean of almost two reinforcements per 10-minute period.

Contingent Positive Reinforcement - Verbal Only. This category includes only verbal reinforcement and again, does not include KOR. In this category DARCEE is highest with a mean of more than three reinforcements per 10-minute period.

All Positive Verbal Reinforcement. Under this category KOR is included. The amounts shown in Table 5 therefore represent all verbal reinforcements, both KOR and contingent. Bereiter-Engelmann is highest and Traditional is lowest.

KOR Positive - Verbal. Only verbal confirmation as to the correctness of response is included in this category. With approximately 13 instances of positive verbal feedback per 10-minutes, the Bereiter-Engelmann program is clearly high.

TABLE 5

Means for Video-Tape Monitoring of Teachers' Behavior

	Programs			
	Bereiter-Engelmann	DARCEE	Montessori	Traditional
Reinforcement				
Contingent Positive	1.83	1.69	0.70	0.51
Contingent Positive Verbal	2.70	3.35	1.31	1.00
Contingent Negative Verbal	0.21	0.09	0.09	0.40
KOR Positive Verbal	13.18	9.06	4.81	3.76
All Negative Verbal[a]	0.82	0.18	0.17	0.66
All Positive Verbal	7.94	6.21	3.06	2.38
Asking				
Academic Verbal Performance	39.42	15.01	5.12	8.89
Academic Non-Verbal Performance	4.04	4.92	6.62	3.85
Content	0.81	4.35	2.56	2.70
Conduct Modification	0.71	1.14	1.32	1.80
All Direct Requests[b]	5.67	3.35	2.29	2.44
Giving				
Academic - Modeling	6.25	0.26	0.37	0.14
Academic - Verbal	6.48	9.10	10.00	10.95
Academic - Manipulatory	0.23	0.18	5.46	0.17
Setting Standards	0.78	0.50	0.59	0.28
Conversing	0.12	0.62	0.34	0.67

[a] Includes KOR.

[b] Includes clarification, opinion, generalization, procedural information, content information, imitation, academic verbal performance, academic non-verbal performance, and conduct.

Note.- Mean frequency for 10-minute period.

Contingent Negative Reinforcement. All negative reinforcement was verbal. This category then includes all negative reinforcement, except knowledge-of-results, whether given for academic or other behavior. Negative reinforcement was infrequent in all programs, but highest in Traditional. Since conduct regulation was high also in Traditional, this high negative contingent reinforcement was apparently given primarily for unacceptable behavior.

All Negative Verbal Reinforcement. This category includes KOR, thus collecting all instances of negative feedback. Here the high program is Bereiter-Engelmann, indicating that most of the negative reinforcement in the Bereiter-Engelmann program was given as KOR for incorrect academic responses.

Direct Requests. In a sense this category represents the number of demands on children since it collects all instances of teachers requesting children to do something, whether academic or otherwise. It is instructive to compare these results with the ratios of teacher activity from the in-class tally sheet. All categories are included from the "Asking" tally sheet, with Bereiter-Engelmann highest and Montessori lowest.

Requests for Conduct Modification. This category includes indirect requests and commands, but consists mostly of direct requests. It is, therefore, the row under all columns for conduct. Traditional was high, indicating that in this program more of the teachers' time was devoted to managing the children's behavior.

Requests for Academic Performance. Requests for academic verbal performance are requests for the children to practice language skills per se. It is clear that such requests occurred most frequently in the Bereiter-Engelmann program, where children received more than 39 such requests on the average in a 10-minute period.

In contrast, requests for academic non-verbal performance and requests for content information did not differentiate greatly among programs, though Montessori was highest in asking for non-verbal performance and DARCEE high in asking for content.

Giving Academic Information. "Academic" on the "Giving" sheet collects three categories which were differentiated on the "Asking" sheet: academic verbal, academic non-verbal and content. With respect to the manner in which information was conveyed to children, comparison of the first three rows and columns under the "Giving" section of Table 5 reveals a number of

interesting facts. Bereiter-Engelmann teachers were high in modeling academic, and they used modeling and direct language instruction about equally. However, their use of language is lower than that of teachers in other programs. Montessori teachers gave academic information through manipulation more than teachers in the other three programs. However, they used direct language instruction also, and twice as often as they used manipulation. Traditional teachers gave academic information almost exclusively by verbal means, despite the de-emphasis on language as a method of instruction in this program. DARCEE teachers also used almost no technique except verbal to give academic information.

All modeling coded was academic in all programs.

Setting Standards. Bereiter-Engelmann was highest and Traditional lowest in setting standards. Very little occurred in any program--the highest amount being less than once per 10-minute period.

Conversing. Results from video-tape monitoring of this category confirms results from the in-class procedure in that DARCEE and Traditional were highest.

2. Grouping Patterns

Table 6 shows a number of aspects of grouping in the four programs.

Number of Groups.

The size of any classroom limits the spatial separation of children and thus restricts the number of groups to the number of locations in which children can be physically isolated. Even though all 20 children were working alone, some would have to be at a table or in a corner of the room in physical proximity. The possible range of number of groups is therefore very narrow. The actual maximum for any class at any time was six. A chi square was calculated on the frequency of occurrence of more than three groups versus three or less. The statistic was significant at the .001 level. Mean frequencies are shown in Table 6 and indicate that more than three groups occurred most often in Montessori and Traditional.

Shifts in Group Size.

Changes in group size were computed as a ratio of changes to number of 15-second tally periods in order to eliminate the effects of number of groups on changes in size. Table 6 shows that shifts in composition of groups

TABLE 6

Grouping Patterns and Total Activity

	Programs			
	Bereiter-Engelmann	DARCEE	Montessori	Traditional
Number of Groups				
Mean frequency of ≥ 3 per 2-hour observation	1.60	.81	3.20	2.35
Shifts in Group Size				
Number per 15-second tally periods	.080	.079	.297	.286
Type of Group Activity Per Tally Period	%	%	%	%
D/NF (Different/no group focus)	.058	.055	.177	.161
S (Same)	.877	.876	.779	.742
D/C (Different/common group focus)	.065	.068	.042	.096
Total Activity				
All acts per 15-second tally periods	.5410	.4309	.3045	.2922

were significantly less frequent in Bereiter-Engelmann and DARCEE than in Traditional. Group size changed as much in Montessori, however, as in Traditional.

Kinds of Groups.

Groups of children who were all engaged in the same kind of activity (S) were significantly more frequent in Bereiter-Engelmann and DARCEE than in Traditional (Table 6). The converse of this is seen under D/NF where it is clear that groups consisting of children who were simply in physical proximity to each other but doing different things were significantly more frequent in Montessori and Traditional. Percentages under the D/C column reveal that there was very little cooperative effort toward a common goal or integrative play among these four-year-olds. In Traditional classes, the absolute amount was slightly greater but not statistically significant.

Total Activity.

Teachers in Bereiter-Engelmann and DARCEE were most actively engaged in teaching, while Traditional and Montessori teachers were less obtrusive in the classroom. Table 6 shows for each program the ratio of teaching techniques of any kind to the number of times tallying was done. Bereiter-Engelmann and DARCEE teachers had higher ratios, indicating more overt teaching than in Traditional and Montessori.

3. Intercorrelations

a. In-Class Monitoring

Intercorrelations on the in-class monitoring variables are given in Table 7. These correlations were based on all 14 teachers in the experimental programs. Since four experimental programs were involved, interpretation of these correlations must be mainly in terms of these programs, rather than in terms of general relationships among such teaching techniques in "typical" classroom situations.

Cell/Tally with Cell/Row Measures

Correlations between the corresponding cell/tally and cell/row measures for each category were significant ($r = .75$ to $.99$), except that for Verbal Instruction (IV) which was .08. This lack of relationship primarily reflected the rank of the Bereiter-Engelmann program on these two scoring procedures for IV. In absolute terms (cell/tally), Bereiter-Engelmann was high on IV, but in relative terms (cell/row), it was low.

TABLE 7

Correlations Among In-Class Monitoring Variables

Teacher	MANP	I-V	X	CON	MANP	I-V	X	MO	RP	CON	Cell/Tally-Cell/Row
[illegible]	[illegible]	[illegible]	[illegible]	4	[illegible]	[illegible]	[illegible]	-39	61	24	75
I-V	-22	[illegible]	-61	11	12	[illegible]	44	-15	-17	19	08
X	-10	71	[illegible]	-57	-32	62	-06	34	-24	-45	84
CON	20	31	-03	[illegible]	34	-68	-46	-16	55	66	81
Children											
MANP	64	-54	-35	43	[illegible]	-62	-31	-46	05	22	86
I-V	-23	56	76	-33	-37	[illegible]	40	15	-63	-44	96
X	-40	02	-11	-42	-34	20	[illegible]	-41	-41	-31	97
MO	-16	16	29	-03	-25	-02	-22	[illegible]	-03	-15	87
RP	41	-55	-37	26	32	-57	-35	25	[illegible]	31	99
CON	55	-21	-23	55	56	-40	-48	-00	55	[illegible]	87

Notes:- N = 14. Cell/Row correlations above diagonal; Cell/Tally below. Correlations significant at $p < .05$ are underlined.

Cell/Tally Measures (Absolute Amount)

For teachers, the only significant correlation among the cell/tally variables was that between IV and Exemplary (X) (r = .71). The ranks of the four programs were identical on these two measures.

For children, cell/tally measures indicated that Conversation (CON) correlated positively with Manipulation (MANP) and Role-Playing (RP) (r = .55), but negatively with X (r = -.48). There was also a negative relation between RP and IV (r = -.57). These correlations reflect the ordering of Montessori and Traditional programs on these variables with Montessori and Traditional being high on CON, MANP and RP, but low on IV and X.

Correlations between teachers and children on the same variables were positive for MANP, IV, and CON. However, there was no relationship for X, reflecting that for teachers, Bereiter-Engelmann was high and Traditional was low; but for children, DARCEE was high and Montessori was low.

The highest correlation for the other teacher-child combinations was between Child-IV and Teacher-X (r = .76); programs ranked the same on these variables. Other positive correlations were between Child-CON and Teacher-MANP and and between Teacher-CON and Child-MANP (both r = .55); with programs high on CON (Montessori and Traditional) being high on MANP for both teachers and children. Negative correlations occurred with Teacher-IV for Child-MANP and Child-RP (both r = -.54); with programs high on Teacher-IV (Bereiter-Engelmann and DARCEE) being low on Child-MANP and RP.

Cell/Row Measures (Relative Amount)

For teachers, in contrast with the positive correlation between X and IV for cell/tally, cell/row measures indicated a negative correlation (r = -.81) between X and IV. As mentioned before, this was because Bereiter-Engelmann was lowest on IV but highest on X for cell/row. CON correlated positively with MANP (r = .44) and negatively with X (r = -.57), again reflecting different emphases in the Bereiter-Engelmann and DARCEE programs as opposed to Montessori and Traditional.

For children, MANP correlated negatively with IV and Motor Activity (MO) (r = -.62 and -.46, respectively), reflecting primarily that Bereiter-Engelmann was low on MANP but high on IV and MO, while Montessori was high on MANP but low on IV and MO. IV also correlated negatively with RP and CON (r = -.63 and -.44, respectively), reflecting a low position for Bereiter-Engelmann on RP and

CON with Traditional high on RP and CON but low on IV.

Correlations between teachers and children on the same variables were positive for MANP and CON (r = .49 and .66, respectively). However, there was no relation between child and teacher measures for IV and X. For IV, DARCEE teachers were high and Bereiter-Engelmann teachers low, yet children in Bereiter-Engelmann classes were high. For X, DARCEE teachers were low but children in DARCEE classes were high.

On other teacher-child combinations, Child-IV, X, and RP were significantly related to teacher-MANP and CON. Child-IV and X correlated negatively with teacher-MANP and teacher-CON (r = -.46 to -.68) while child-RP correlated positively with the same teacher variables (r = .61 and .55, respectively). On these variables, Montessori and Traditional teachers were high on MANP and CON while children in these programs were low on IV and X, and high on RP. Other positive correlations were between teacher-IV and child-X (r = .44) and between teacher-X and child-IV (r = .62). However, teacher-X correlated negatively with child-CON (r = -.45), with Bereiter-Engelmann teachers high and Bereiter-Engelmann children low.

Summary

Correlations between the absolute and relative measures of the same categories were all positive and high except for IV. The lack of relationship on IV was due to the high position of Bereiter-Engelmann on cell/tally and its low position on cell/row.

A larger number of significant correlations occurred among the cell/row measures than among the cell/tally measures, primarily because the scoring procedure for cell/row made each category dependent upon each of the others. This intrinsic dependency among the cell/row measures makes interpretation of the correlations difficult. In general, correlations among both the absolute and relative measures reflect the different emphases between the more and less didactic programs.

b. Video-Tape Monitoring

No interpretation of the video-tape correlations have been made as yet.

Summary of Treatment Dimensions

Evaluation by consultants indicates that the implementations of the four programs were average. Results from monitoring of classes are more convincing. They indicate that most of the characteristics of the programs which it was feasible to assess in this manner were present to a sufficient extent to produce greater homogeneity within programs than among them. The four programs were, in any case, demonstrably different along a number of important dimensions.

Descriptively, in Bereiter-Engelmann classes, teachers did a great deal of verbal instruction, exemplification and modeling, provided large amounts of feedback--both positive and negative, elicited significant amounts of verbal recitation from the children and did a great deal of modeling. Children in these classes did relatively little manipulation of materials, role-playing, or conversing with each other or the teacher. What they did most was verbal recitation.

DARCEE teachers also used rather large amounts of verbal instruction, and relative to other techniques had more conversation with the children than was the case in Bereiter-Engelmann. They were second only to Bereiter-Engelmann in positive feedback. DARCEE children did more verbal recitation than anything else.

Montessori teachers were low in most of the techniques tallied, which is consistent with the unobtrusive role which Montessori teachers assume in the classroom. They gave little reinforcement or KOR. Their children were significantly high in manipulation which is also consistent with the program since it leans very heavily on the use of materials designed to teach through the child's manipulation of them.

Traditional teachers used manipulation of materials more than any other technique and had more conversation with their children than Bereiter-Engelmann teachers. Children in Traditional classes were significantly higher in role-playing both absolutely and relatively than in the other three programs. Traditional teachers provided little positive feedback compared to the other programs, except that they used negative reinforcement to a greater extent than any other program, most of this being directed toward behavior control rather than negative feedback for errors in academic performance.

Academic information was given about equally in all programs, but the manner in which it was given differentiated programs. Bereiter-Engelmann teachers modeled and used

verbal instruction, Montessori teachers used manipulation and verbal instruction, while Traditional and DARCEE teachers gave information almost exclusively through verbal instruction.

D. Teacher Characteristics

Personality

Personality variables in teachers as assessed by the 16 Personality Factors did not differentiate programs. A Kruskal-Wallis one-way analysis of variance was used to compare programs on the four summary factors provided by the test: low anxiety-high anxiety, introversion-extraversion, tenderminded emotionality-alert poise, and subduedness-independence. Since there were no significant differences among programs on these four summary factors, no analyses were made on the smaller factors making up the summaries.

Intelligence

Teachers' IQ as assessed by the Peabody Picture Vocabulary Test were also compared by means of the Kruskal-Wallis one-way analysis of variance. Differences were not significant although the two Montessori teachers had IQs above those of teachers in any other program, and the statistic approached significance at the .05 level.

Attitudes toward Teaching

Neill's Questionnaire for teachers was designed to assess attitudes in five areas: attitudes toward teaching, toward knowledge, toward peers, toward self, and toward pupils. Programs were compared by means of Kruskal-Wallis analysis of variance on each of these factors and on the total score. No significant differences were found. In general, teachers' scores corresponded to the levels reported by Neill in a standardization group which also consisted of Head Start teachers in Louisville.

Agreement with Program Philosophy

Results on the Statements Test designed to determine the extent of agreement by teachers with statements taken from their own and the other three programs are shown in Table 8. No statistical analyses were made but it is clear that mean agreement scores are not outstandingly high with their own program for any group of teachers with the exception of Montessori. In some cases agreement scores were higher with statements taken from other program materials than from the program in which the teacher was trained. The evidence from monitoring of classroom behavior, however, indicates a high degree of congruence between

TABLE 8

Teachers' Agreement with Statements from Program References

	Source of Statement											
	Bereiter-Engelmann			DARCEE			Montessori			Traditional		
Teachers	Ag.	Dis.	A-D	Ag.	Dis.	A-D	Ag.	Dis.	A-D	Ag.	Dis.	A-D
Bereiter-Engelmann	55	28	37	58	9	49	54	12	42	38	36	2
DARCEE	46	19	27	59	8	51	61	8	53	27	32	-5
Montessori	42	22	20	41	17	24	79	8	71	46	21	25
Traditional	38	32	6	59	15	44	48	21	27	59	20	39

Note.- All figures represent percentages based on "Statements Test".

program philosophy and what the teachers were actually doing. It might be concluded that the Statements Test is not a valid measure of attitudes toward program philosophy, but some degree of validity is indicated by the responses of the consultants for the four programs which are shown in Table 9. Consultants obviously agreed more often with the statements taken from their own program materials. This is particularly true in Bereiter-Engelmann and DARCEE programs. Disagreement scores in their own program statements are also quite low for the consultants as compared with the teachers.

To some extent the teachers' lower scores may simply be due to a lesser degree of familiarity with the materials and to greater difficulty in interpreting the meaning of the statements, some of which even the consultants found ambiguous. On the other hand, the discrepancy between attitudes expressed on a questionnaire and actual behavior in a classroom underscores the necessity for objective assessment of classroom activity. It is also interesting that in each of the four programs the teacher who was rated best by the consultant had the highest agreement score within that program, although consulants' choices were not consistent with class order on relevant dependent variables.

Summary

Results on all measures of teacher differences which were assessed indicated that there were no significant differences among programs. Teachers in the four programs did not differ with respect to the measures used to assess personality characteristics, IQ, attitudes toward teaching, nor in attitudes toward programs. Although there were substantial differences among individual teachers, these differences, insofar as we were able to measure them, do not seem to have been a source of program effects.

E. Treatment Effects

Across Programs

1. Method of Analysis

Analysis of covariance with the fall measure as the covariate was planned for the analysis of all variables on which there were both fall and spring measures. However, examination of the fall means indicated that some differences between programs might exist. Analysis of variance on the fall measures did, in fact, indicate significant differences among programs for some variables. For example, differences among programs occurred on the Stanford-Binet with the programs ordering as predicted. Such program differences, the similar

TABLE 9

Consultants' Agreement with Statements from Program References

	Source of Statement											
	Bereiter-Engelmann			DARCEE			Montessori			Traditional		
Teachers	Ag.	Dis.	A-D	Ag.	Dis.	A-D	Ag.	Dis.	A-D	Ag.	Dis.	A-D
Bereiter-Engelmann	81	9	72	77	13	64	46	46	0	17	74	-57
DARCEE	52	21	31	92	0	92	50	27	23	17	63	-46
Montessori	49	23	26	62	17	45	77	0	77	57	20	37
Traditional[a]	38	39	-1	66	17	49	62	14	48	68	10	58

Note.- All figures represent percentages based on "Statements Test".

[a]Mean of two.

ordering of classes within programs on related variables, and the fact that subjects had been randomly assigned to programs indicated that programs were already having some effect. These differences were unexpected, although the final test (Stanford-Binet) was administered after approximately 8 to 10 weeks of school, equivalent in time to some summer Head Start programs. It seemed more appropriate to view the fall and spring measures as two points in time, rather than as pre- and post-tests.

Analysis of covariance, therefore, was inappropriate because the covariate was not independent of the experimental treatment. Elashoff (1969) and Sprott (1970) state that violation of this assumption is serious and cannot be overcome or reduced by transformation of data, random assignment of subjects, normality of distribution of the covariate, or cautious interpretation of the covariance analysis.

Repeated measures analysis of variance was therefore used for the dependent variables. Such a procedure yields more precision, power, and information than analysis of change scores only. Through use of multiple comparison techniques, programs could be evaluated on the basis of their ordering in the spring as well as the amount of change from fall to spring.

The experimental design was completely replicated for three of the programs (Bereiter-Engelmann, DARCEE and Traditional) in four areas of the city. It was replicated with the four programs, by the addition of Montessori classes, in two of these areas. Several related analyses were necessary in order to adequately assess treatment effects.

In order to compare the four programs and controls, a 5x2x2 (four programs and controls by sex by fall-spring) unweighted means analysis of variance with repeated measures on the last factor was used. This analysis will be referred to as the one-way analysis. In order to further compare programs and teachers (classes within programs), area was introduced as an additional variable for two other repeated measures analyses of variance: (1) a 3x4 analysis which compared the three programs in all four areas with sex and fall-spring as factors (3x4x2x2) and (2) a 4x2 analysis which compared all four programs in the two areas with sex and fall-spring as factors (4x2x2x2). In both of these analyses, it was necessary to equalize the number of males and females within a given class. To achieve this equality, subjects were randomly pulled from each class. These two analyses provided increased precision over the one-way because the variance due to individual classes (teachers) could be assessed with area-by-program interactions.

Obviously these three analyses were dependent. All significant effects on the one-way analysis are reported. Because of increased precision, the 3x4 and the 4x2 analyses sometimes revealed program main effects and/or interactions not found on the one-way. For the 4x2 only significant effects and/or interactions which involved Montessori are reported. For the 3x4 any additional program effects and/or interactions are reported.

The Scheffe Test (Scheffe, 1953) was used to compare means. It allows contrasts involving combinations of means, controls the alpha level for all possible contrasts and is the most conservative multiple comparison test for pairs of means (Kirk, 1968; Winer, 1962). A conservative test was desired because of the relationships between some of the dependent variables. In addition, many contrasts which examined degree of change and sex involved combinations of several means.

Calculation of the appropriate Scheffe coefficient was made according to Levin and Marascuilo (1970). They distinguished between two types of post hoc comparisons which can be made in examining significant interactions: those involving only comparisons between cell means at one level of a factor and those involving the joint effect of the various levels of the factors. Different standard errors are used in the Scheffe coefficient for these two contrasts, resulting in a conservative test for comparing cell means. In the present study, both types of comparisons were of interest. For example, for significant fall-spring by program interactions, different amounts of change for programs as well as program ordering in the spring were examined (cell means). The .05 level was used for all statistical tests.

2. Cognitive Measures

Table 10 presents fall and spring means for the main test battery. Two of the three instruments used to assess cognitive development, the Stanford-Binet (Form L-M) and the Preschool Inventory, showed very similar results with respect to program differences. The product-moment correlation between the Stanford-Binet and the Preschool Inventory for the entire group was .61. On the Binet there was a program effect over both fall and spring testing. The Bereiter-Engelmann and DARCEE programs were higher than controls. On the spring test only, Bereiter-Engelmann and DARCEE were significantly higher than controls and in addition Bereiter-Engelmann was significantly higher than Traditional.

Figure 9 reveals a puzzling result with respect to DARCEE children in that their mean was significantly high on the fall test but did not change between fall and spring. If fall test levels are interpreted as representing program effects at the end of eight weeks, this result suggests that the DARCEE program produced rapid gains but did not continue to affect IQ at the same rate over the succeeding 6-month period.

TABLE 10

Fall-Spring Means on Main Test Battery for Programs and Controls

	Tests								
	Fall 1968								
Program	S-B	PSI	Q	C-A	C-V	R-P	R-R	D-B	EBF
Bereiter-Engelmann	93.25	26.33	12.47	17.72	1.98	21.21	9.73	3.21	7.81
DARCEE	96.02	28.92	13.28	14.98	2.20	20.41	8.35	3.58	7.03
Montessori	91.50	25.21	11.88	19.76	1.58	19.72	9.62	4.06	7.79
Traditional	89.35	24.36	11.87	17.32	1.92	20.79	8.67	3.30	7.45
Controls	89.21	28.29	12.21	16.59	3.94	17.47	8.18	4.06	7.29
	Spring 1969								
Bereiter-Engelmann	99.52	39.06	13.88	18.06	1.05	22.13	9.48	4.19	9.77
DARCEE	97.56	40.98	13.30	17.81	2.17	22.75	10.10	6.36	10.20
Montessori	96.34	37.55	13.79	18.67	1.58	22.07	7.72	5.61	9.52
Traditional	95.02	35.98	13.87	17.15	1.21	22.35	8.96	4.23	10.19
Controls	90.00	33.18	13.74	14.09	1.32	20.65	7.06	4.97	9.53

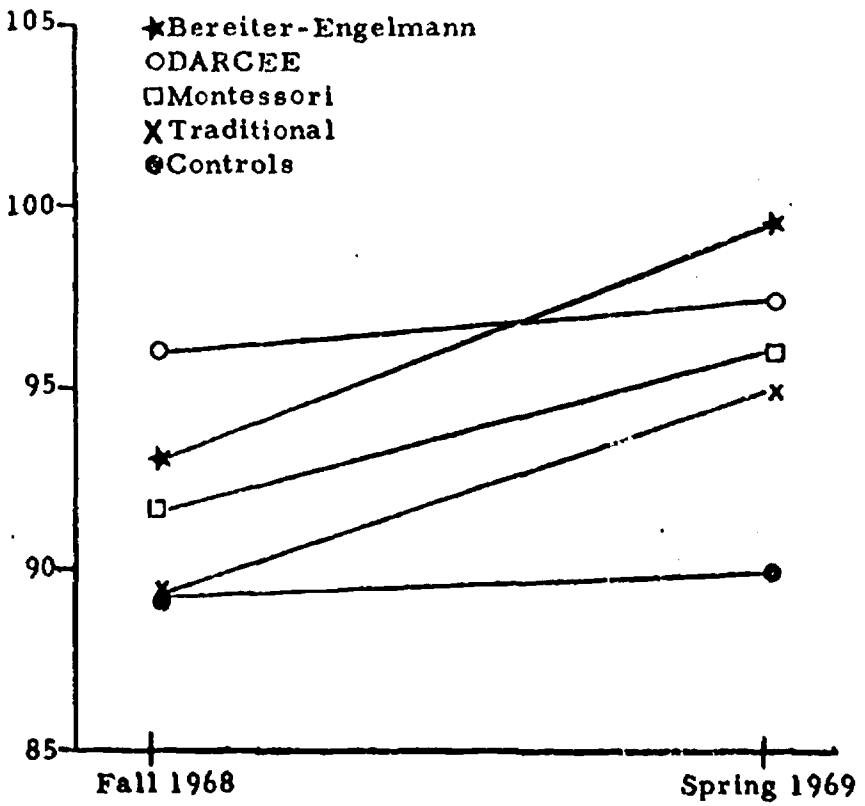

Fig. 9. Fall-Spring Means on Stanford-Binet for programs and controls.

In order to determine whether Binet IQ changes were concentrated at lower levels of IQ, a cumulative plot for controls and experimentals on the second test was made. This distribution is shown in Figure 10. This graph demonstrates several interesting facts. First, IQs appear to be normally distributed with a fairly wide range. Second, superiority of experimental subjects was not confined to any particular IQ level but appears rather evenly distributed over the entire range.[1]

On the Preschool Inventory (Figure 11), experimental programs gained significantly more than controls, and on the spring test the DARCEE, Bereiter-Engelmann, and Montessori children scored significantly higher than controls. DARCEE children were higher than children in the Traditional program.

Thus the results on the Stanford-Binet and on the Preschool Inventory are very similar.

Results on the Quick Picture Vocabulary are not so congruent with those of the other two cognitive measures. However, for this sample the Quick correlated only .42 with the Binet and .43 with the Preschool Inventory. On this test Figure 12 shows that all programs improved about the same with the exception of the DARCEE program in which there was virtually no change from first to second testing. This result is similar to the Binet results for the DARCEE program, except that on the Quick the other programs changed enough to reach a higher level than DARCEE in the spring test. In view of the fact that the sample of children in the DARCEE program scored highest in the spring on the Peabody Picture Vocabulary Test, this result is difficult to explain. Reliability of the Quick was low, with the test-retest correlation being .39, whereas test-retest correlations for the Binet and PSI were .70 and .79, respectively. In view of all these facts, a tentative conclusion would be that the Quick may not have been a valid measure of cognitive change for this particular population. However, the similarity of results for the DARCEE program to those on the Binet for this program remains unexplained.

Table 11 presents results from several tests which were given only in the spring to a sample of six children from each experimental class. Four of these were given to assess achievement in specific areas. These were: Parallel Sentence Production, Basic Concept Inventory, Arithmetic, and Expressive Vocabulary. The fifth was the Peabody Picture Vocabulary Test which was given in order to compare results on this test with those on the Quick Picture Vocabulary Test. Significant program effects were found on Arithmetic and Parallel Sentence Production. On Arithmetic both Bereiter-Engelmann and DARCEE children scored higher than children in the Traditional program and Bereiter-Engelmann scores were also higher than Montessori.

[1] Control curve has been smoothed slightly to eliminate minor irregularities resulting from the smaller number of subjects.

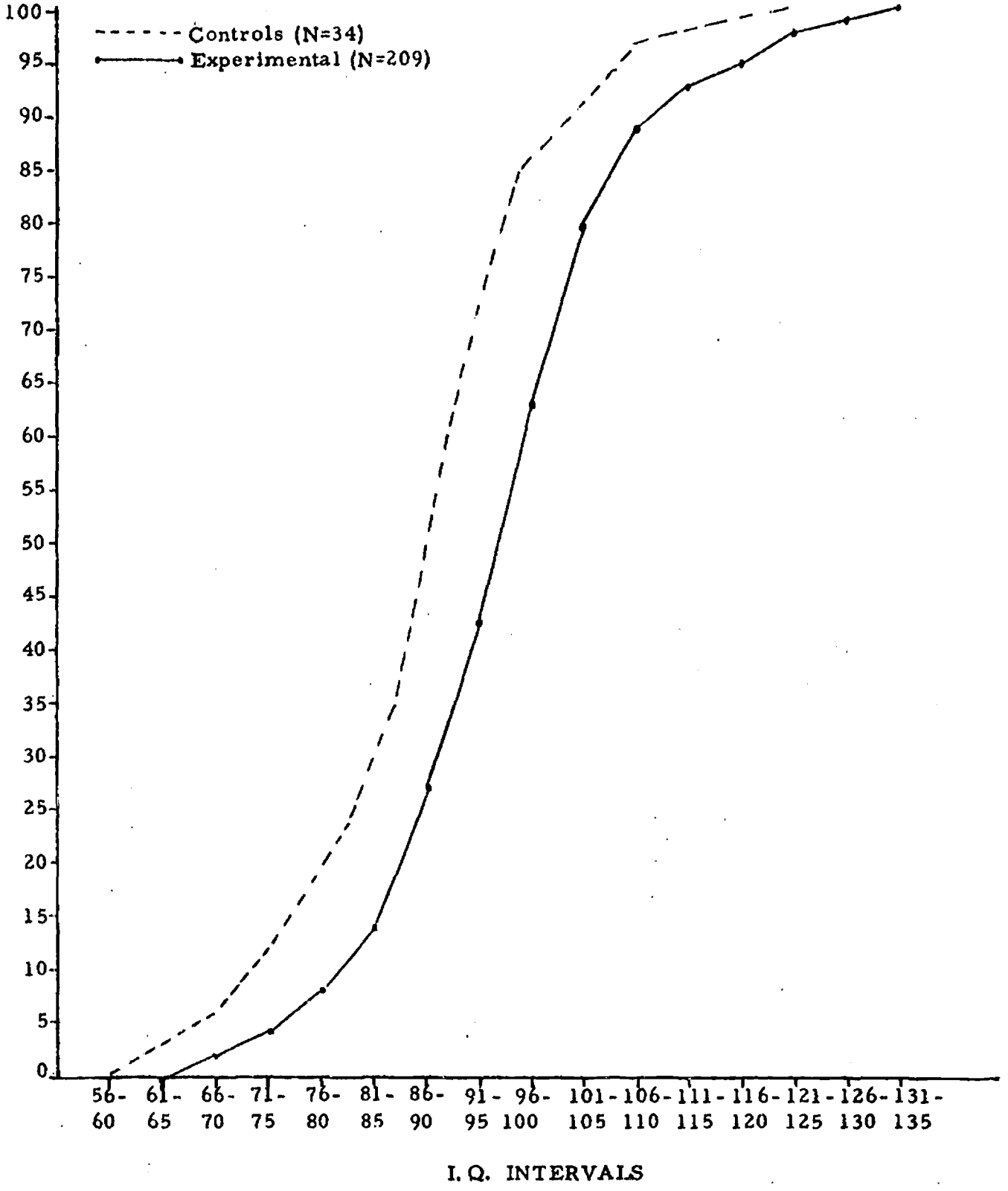

Fig. 10. Cumulative Distribution of STANFORD-BINET I. Q. 's after six months of special programs for experimental Ss and no preschool for controls. (Control curve slightly smoothed.) (Ordinate figures are percentages.)

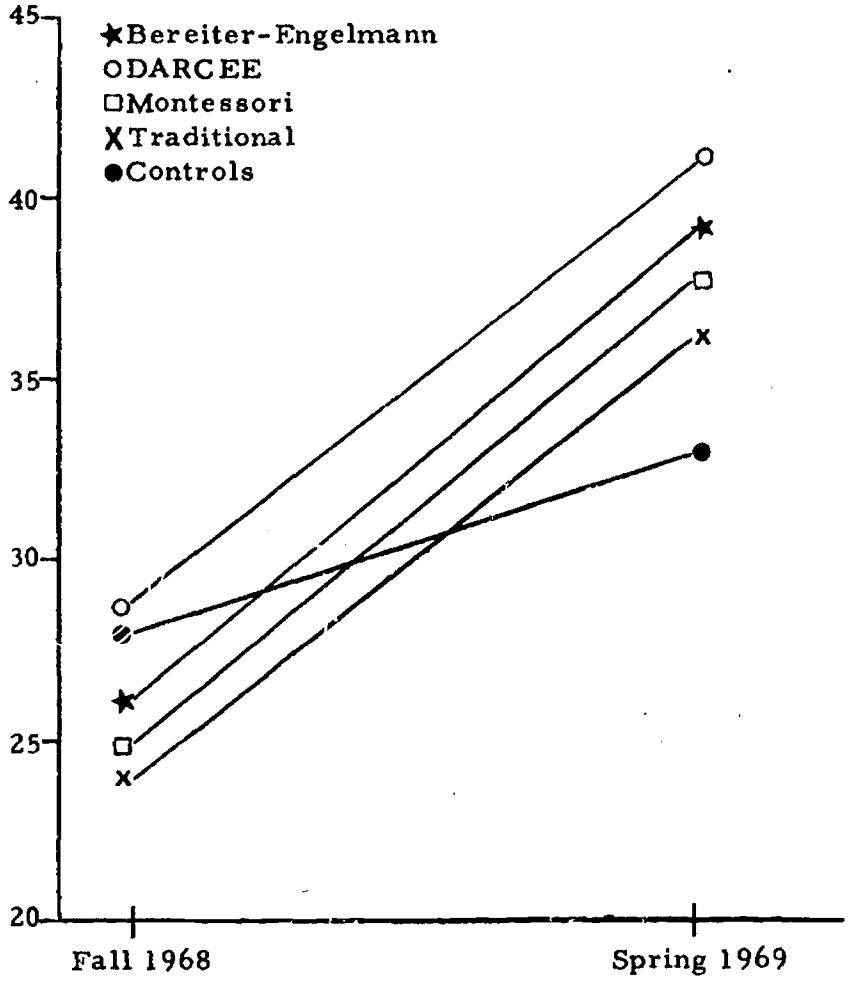

Fig. 11. Fall-Spring means on Preschool Inventory for programs and controls.

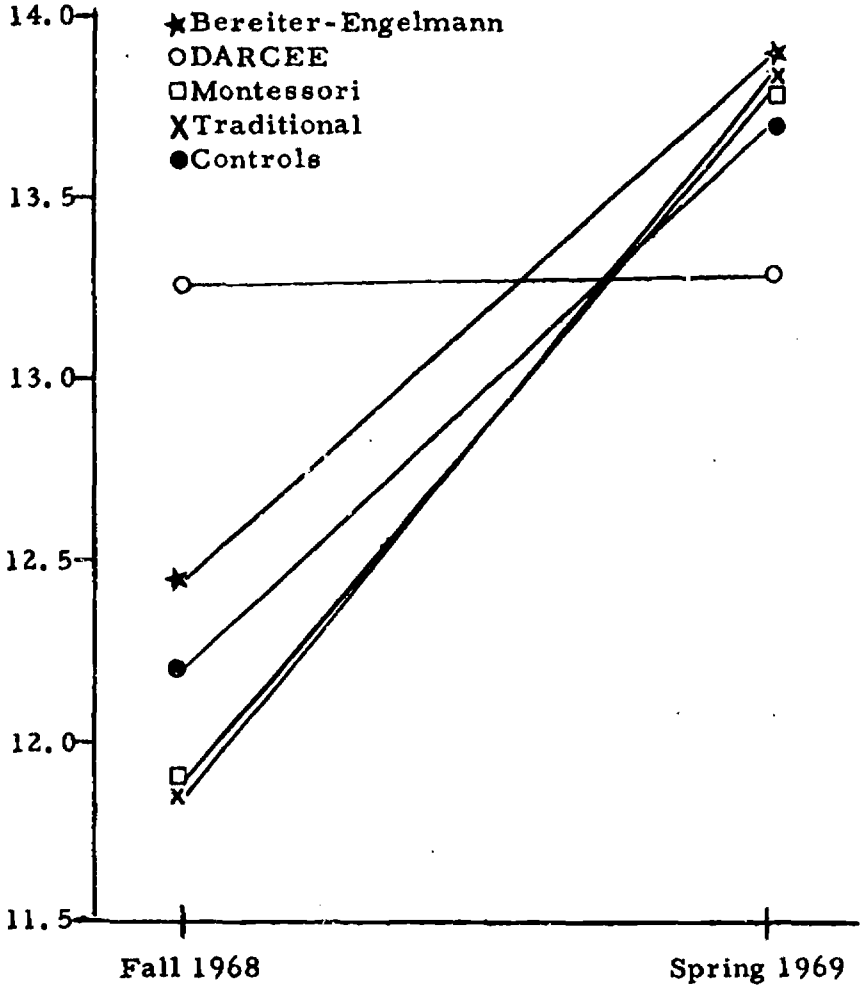

Fig. 12. Fall-Spring means on Quick Picture Vocabulary for programs and controls.

TABLE 11

Means on Additional Spring Tests for Programs

	Tests				
	PPVT	BCI[a]	ARIT.	PSP	EVI
Program					
Bereiter-Engelmann	37.71	36.17	17.75	95.88	27.38
DARCEE	42.42	37.79	13.42	90.58	26.63
Montessori	38.00	35.00	8.47	84.83	27.18
Traditional	37.88	44.54	6.67	77.88	25.08

Note.- N=84

[a] Low score is optimum.

On the Parallel Sentence Production Bereiter-Engelmann was superior to Traditional.

Sex Differences

A main effect of sex appeared on the Stanford-Binet. Figure 13 shows that females were higher than males. Although not statistically significant, two interesting factors may be noted--the large sex difference in Montessori on the spring test, where the females were very close to the highest group (Bereiter-Engelmann) and males were below all other programs; and the very small difference between the sexes in the Traditional program on both testings.

On the PSI, there was a significant sex-by-program interaction which is shown in Figure 14. With controls included the interaction was not quite significant at the .05 level. Figure 14 shows that in DARCEE females were higher than males, in Bereiter-Engelmann there was no difference between the sexes, and in Traditional males were higher than females. Programs apparently did not differentiate males on this test although all programs had significant effects on males from first to second test as shown in Figure 15. Females also gained from first to second test in all programs (Figure 15) but in contrast to males were apparently affected differentially by the four programs. There were no sex effects on the Quick. No analyses by sex were made on the achievement tests because of the small sample size.

Summary

In summary, children in all experimental programs were in some way superior on cognitive measures as compared with the controls, but the best effects were obtained in Bereiter-Engelmann and DARCEE. In addition, these two programs produced significant gains in achievement measures in the areas of numerical and linguistic ability.

Sex effects were found, consisting primarily of superior scores for females on the Stanford-Binet and a program effect for females but not for males on the PSI.

3. Motivational and Social Measures

Motivational and social variables were assessed in three ways: by tests selected from the Cincinnati Autonomy Battery, by ratings made by two groups of testers, and by ratings of children in classes completed by teachers and also by aides.

(a) Tests

Results from the tests selected from the Cincinnati Autonomy Battery are shown in Table 10. In verbal

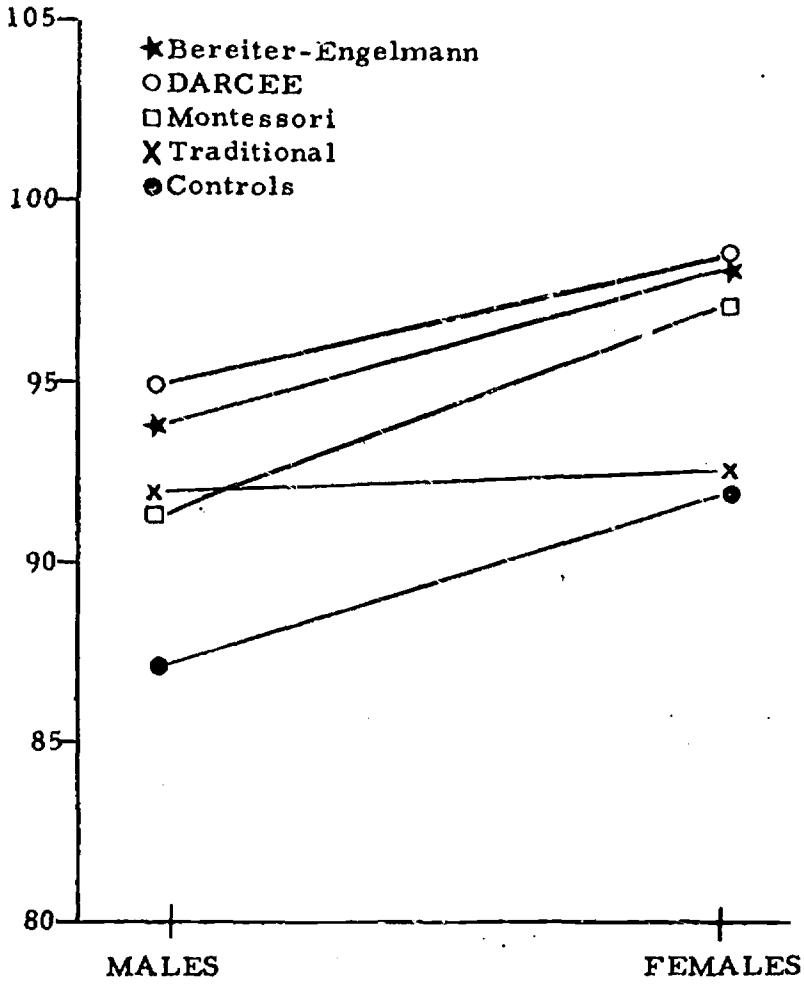

Fig. 13. Male-Female means on Stanford-Binet for programs and controls.

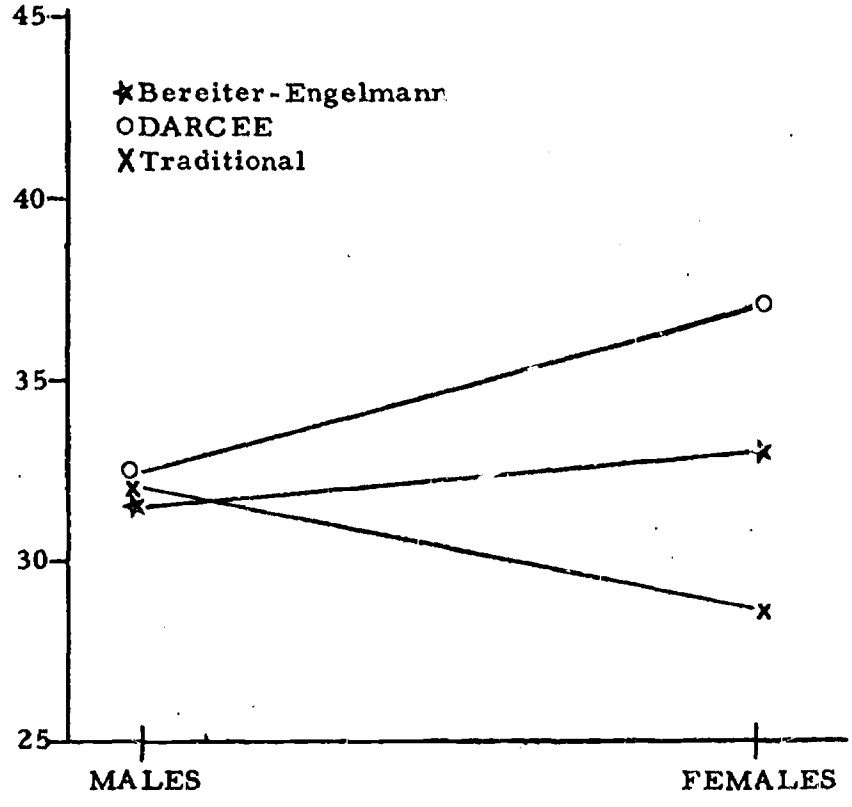

Fig. 14. Male-Female means on Preschool Inventory for Bereiter-Engelmann, DARCEE, and Traditional programs.

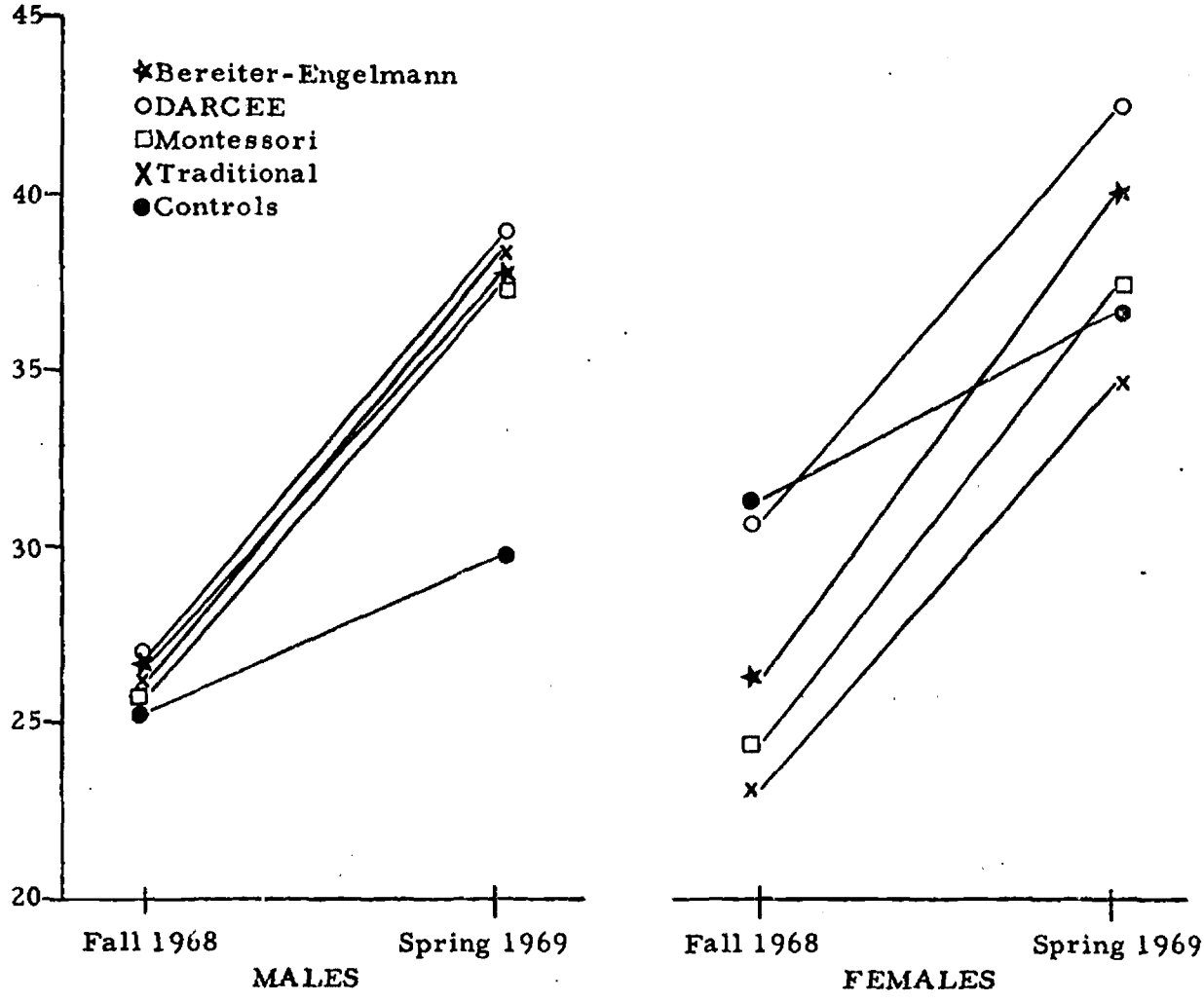

Fig. 15. Fall-Spring means for Males and Females on Preschool Inventory for all programs and controls.

xpression of curiosity, there were no program differences but Figure 16 shows that controls were significantly high on the first test in the fall and decreased rather dramatically toward the end of the year. In curiosity activity Figure 17 shows that DARCEE children gained, children in the other three programs did not change significantly and controls decreased. In task persistence (Figure 18), as measured by the Replacement Puzzle, DARCEE Bereiter-Engelmann and Traditional were superior to controls over both fall and spring tests. All groups, including controls, improved significantly from fall to spring. In resistance to distraction (Figure 19), also measured by the Replacement Puzzle, DARCEE was the only program to gain. Bereiter-Engelmann and Traditional classes did not change while controls and Montessori children decreased. In inventiveness as measured by the Dog and Bone (Figure 20), DARCEE children were significantly higher on the second test than Bereiter-Engelmann and Traditional.

Sex Effects

There were no significant sex effects on any of the motivational tests except on inventiveness (Dog and Bone). On this test a sex interaction occurred (Figure 21) which consisted of a greater gain by males than by females from first to second test. This was consistently the case in all programs, but did not occur in the control group in which there was little improvement for either males or females. Control females, however, were higher than any other group on the first test, and retained their relative position on the second test, being surpassed only by DARCEE males and nearly equalled by DARCEE females.

(b) Ratings

Behavior Inventory

Scores on the Behavior Inventory were adjusted so that a high score was "good" regardless of the name of the factor. In other words a change from lower to higher score on aggression does not mean more aggression--it means an improvement, or less aggression.

Results for ratings by both teachers and aides on the five factors assessed by the scale are shown in Table 12. Analyses of the ratings by aides are not presented. Significant results from teachers' ratings were: Timidity - DARCEE children imporved more than those in other programs. Independence - DARCEE children increased more than those in the other three programs. Verbal-Social Participation - DARCEE and Bereiter-Engelmann gained more, but a program main effect over both ratings indicates that DARCEE children were superior to those in the other

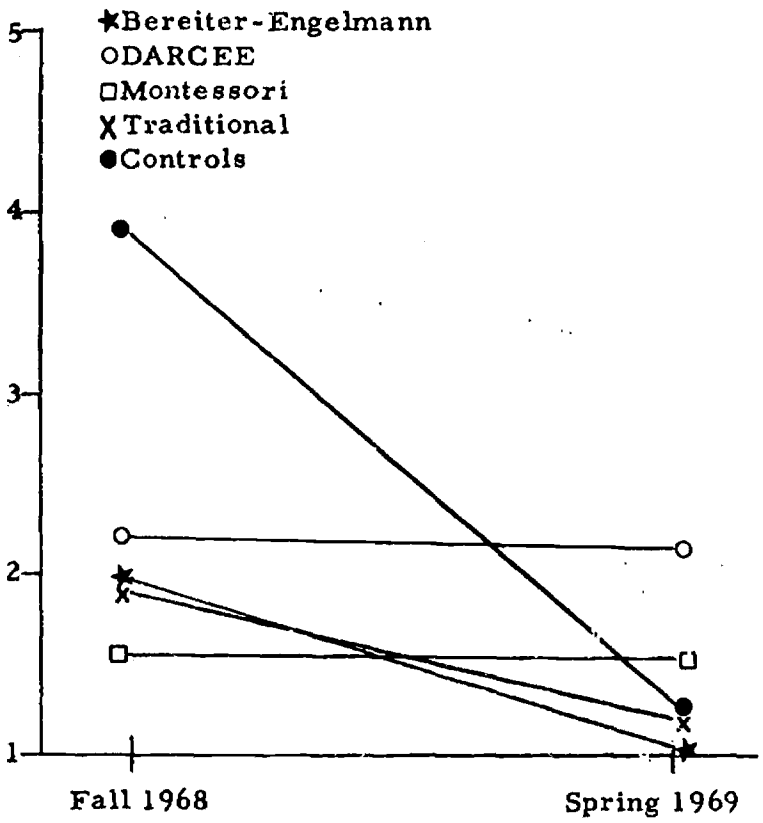

Fig. 16. Fall-Spring means on Curiosity-Verbal for programs and controls.

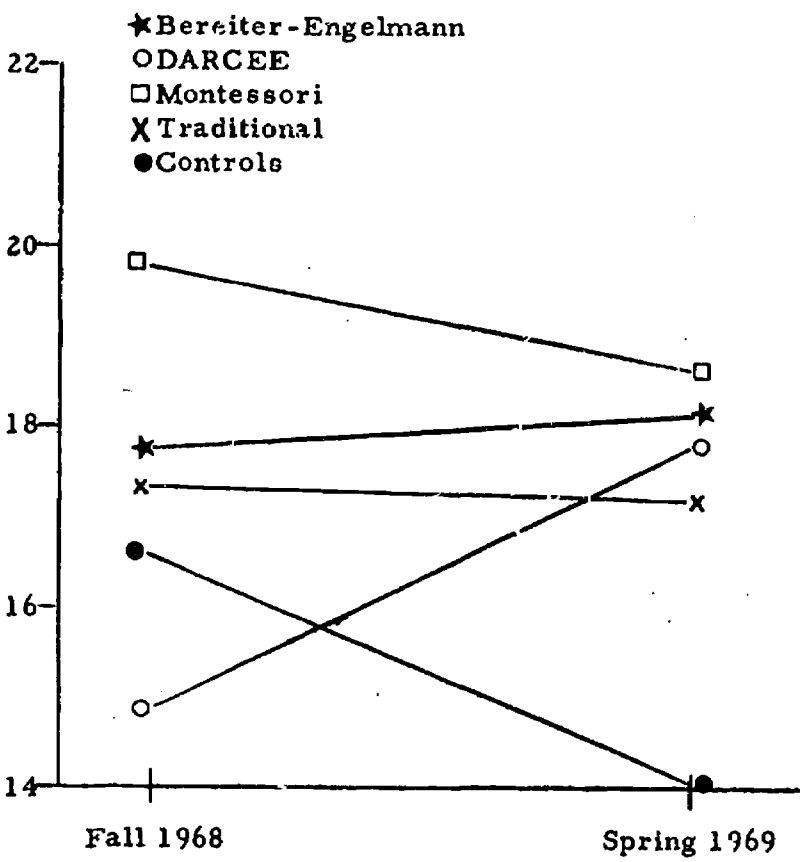

Fig. 17. Fall-Spring means on Curiosity-Activity for programs and controls.

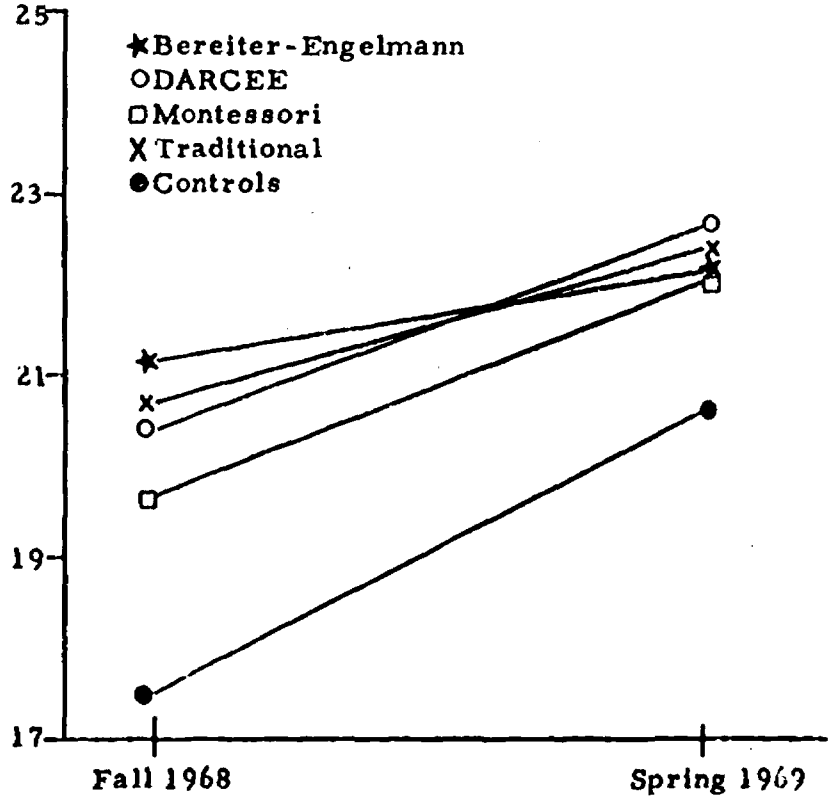

Fig. 18. Fall-Spring means on Persistence (Replacement Puzzle) for programs and controls.

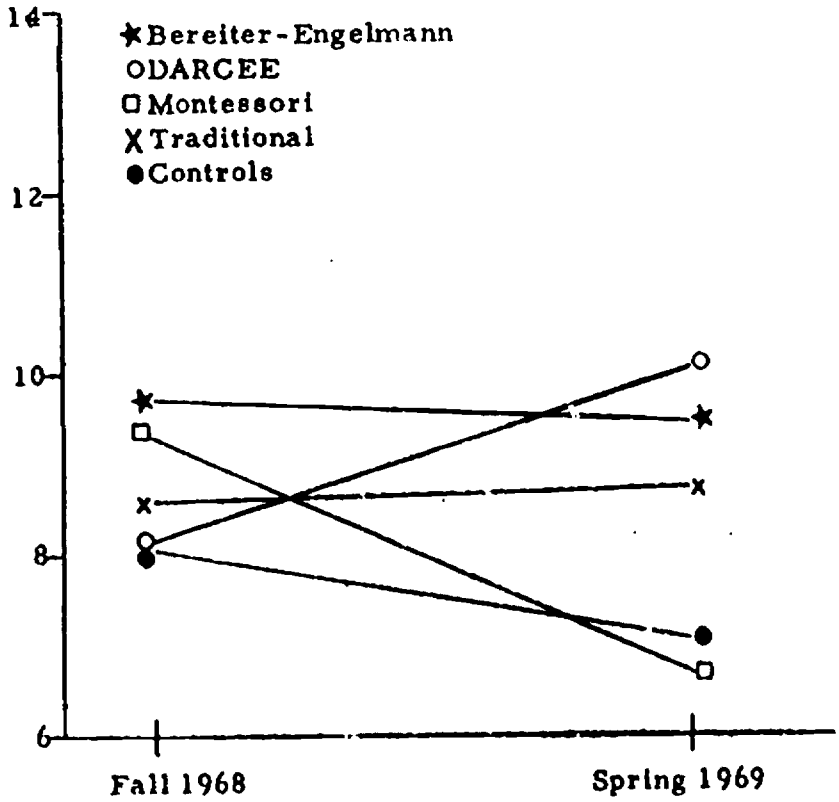

Fig. 19. Fall-Spring means on Resistance to Distraction (Replacement Puzzle) for programs and controls.

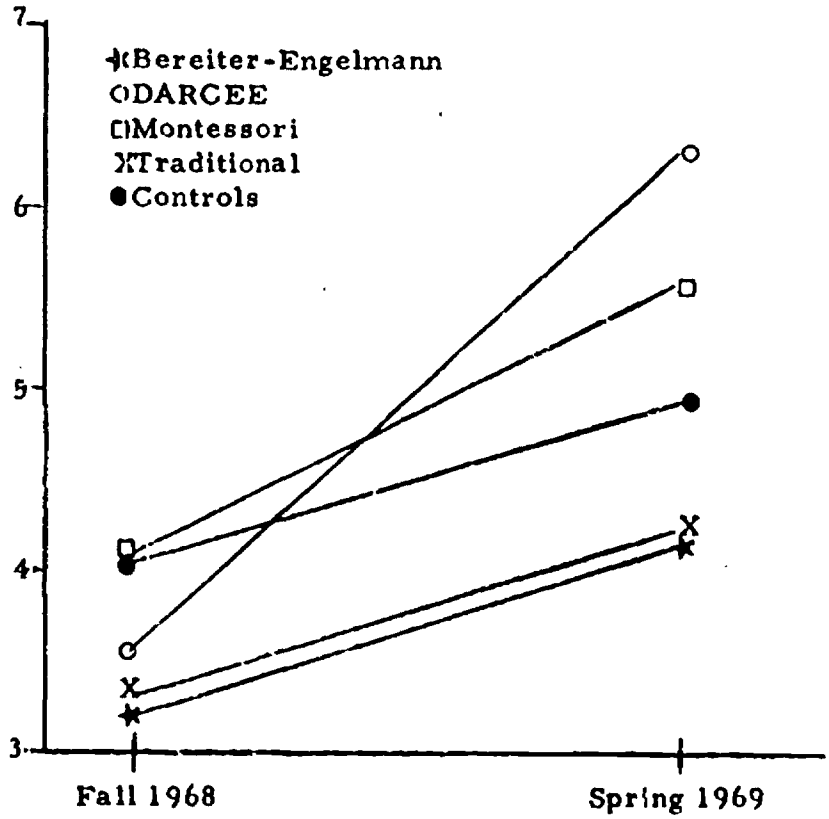

Fig. 20. Fall-Spring means on Inventiveness (Dog and Bone) for programs and controls.

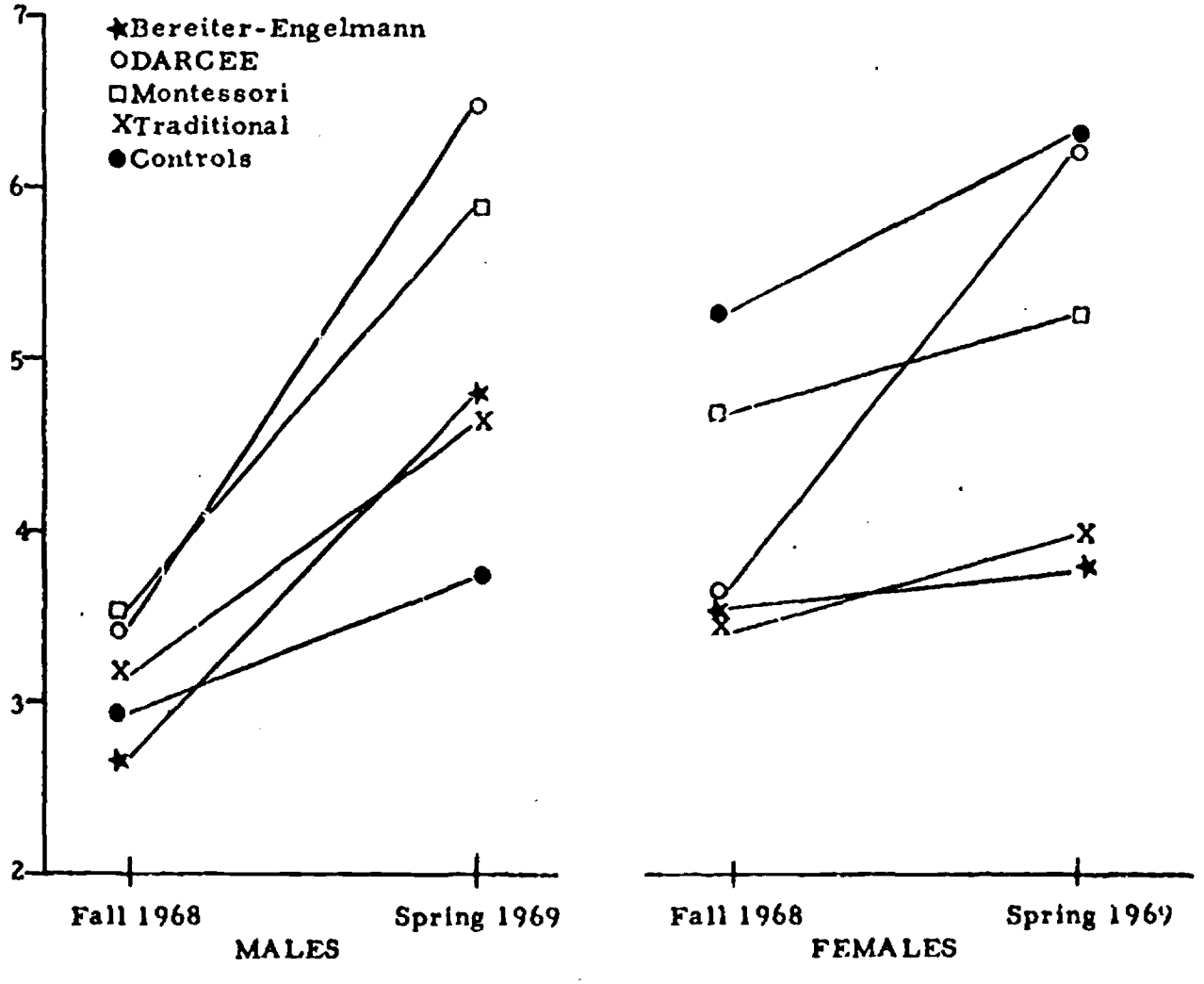

Fig. 21. Fall-Spring means for Males and Females on Inventiveness (Dog and Bone) for programs and controls.

TABLE 12

Fall-Spring Means on Behavior Inventory Ratings by Teachers and Aides

Fall 1968

	Teachers' Ratings					Aides' Ratings				
Program	TM	VP	IN	AG	AC	TM	VP	IN	AG	AC
Bereiter-Engelmann	11.97	10.03	11.73	13.34	11.95	10.88	9.84	10.46	11.55	10.46
DARCEE	11.61	11.51	11.05	12.76	11.92	11.88	11.60	11.38	12.80	12.63
Montessori	12.97	10.41	12.38	12.50	11.28	11.51	11.58	11.14	10.57	10.65
Traditional	12.58	11.83	12.75	13.40	12.65	11.69	11.60	11.72	12.12	11.38
Spring 1969										
Bereiter-Engelmann	12.10	12.07	11.95	13.78	11.85	12.16	12.47	11.18	12.84	11.40
DARCEE	13.97	14.39	13.19	13.20	13.29	13.69	12.40	13.41	13.29	13.89
Montessori	13.06	10.94	12.72	12.66	11.28	12.58	13.37	13.26	13.01	11.96
Traditional	12.54	12.92	13.10	12.21	12.38	11.72	12.74	12.00	11.58	11.43

three programs. Aggression - Traditional children were rated more aggressive in the spring than in the fall while children in all other programs were rated less aggressive in the spring. Bereiter-Engelmann children were rated significantly lower in aggression than Traditional in the spring. Achievement differences are primarily at the teacher rather than the program level, but again DARCEE children were high in the spring ratings.

Sex Effects

There were no sex effects or interactions on the Behavior Inventory. It might be expected that teachers would use different criteria in rating the two sexes, or would be inclined to view similar behavior differently--for example, to see males as more aggressive. Apparently, they did not.

Teacher-Aide Correlations

Correlations between ratings by teachers and ratings by aides on the Behavior Inventory are shown in Table 13. Although the correlations are in general significant, ranging from .42 to .64, they are not as high as one would like between groups of raters who spend equal amounts of time with children. Correlations are highest between teachers and aides. In Bereiter-Engelmann and DARCEE, and this may reflect the grouping of children in these programs which affords both teachers and aides a more systematic situation for observation and insures a more even distribution of attention to all children in the class. The consistently high correlations between teachers and aides in the DARCEE program on the first rating, however, raise a question as to their independence, particularly since the correlations are not so high at the end of the year.

Binet Face Sheet

Both the PSI and the Binet testers rated subjects on the Face Sheet of the Stanford-Binet. However, since complete fall ratings for the Binet testers were not available, complete analyses were made of the Face Sheet ratings fro the PSI testers only.

The Face Sheet was scored for three factors: Achievement motivation, confidence in ability, and activity level. For all subjects, experimental and control, high correlations existed (see Table 15) among these three factors for both the PSI and Binet testers. The pattern was similar for both groups of testers for each experimental program and the control sample. Therefore, there is some doubt as to

TABLE 13

Teacher-Aide Correlations on Behavior Inventory

	Tm.	VSP	Ind	Agg	Ach
All Programs[a]					
Pre	0.56	0.61	0.50	0.42	0.44
Post	0.64	0.47	0.56	0.59	0.62
Program					
Bereiter-Engelmann					
Pre	0.61	0.63	0.39	0.40	0.68
Post	0.71	0.60	0.55	0.69	0.68
DARCEE					
Pre	0.73	0.70	0.73	0.73	0.73
Post	0.49	0.33	0.70	0.56	0.71
Montessori					
Pre	0.30	0.74	0.51	0.22	-0.23
Post	0.47	0.43	0.23	0.75	0.30
Traditional					
Pre	0.51	0.36	0.37	0.37	0.13
Post	0.65	0.56	0.52	0.39	0.56

[a] Correlations significant for all programs at $p < .05$ are underlined.

whether these factors were independent for our sample. The factors are presented separately, however, pending the results from kindergarten testing, where greater differentiation may occur.

Ratings by both groups of testers are shown in Table 14 for all three factors. In general, Binet testers rated all programs higher than did the PSI testers, but the relative positions of programs were similar with the exception of the control group (Figure 22). For this group, the Binet testers' ratings were higher than those of Montessori and Traditional, whereas the PSI testers gave the control group the lowest ratings on all factors as compared with experimental programs.

Analysis of variance on the PSI testers' ratings indicated that for achievement motivation, DARCEE and Bereiter-Engelmann children gained more than the other three groups, and on the spring test these two programs combined were superior to the combination of controls and Montessori. This result is fairly consistent with the teachers' ratings for achievement motivation on the Behavior Inventory, where DARCEE was high and Bereiter-Engelmann second.

Sex Effects

A sex-by-program interaction appeared for the Confidence in Ability Factor and a similar interaction was almost significant for the other two factors. In view of the high correlation among these factors, the three have been combined (Figure 23) in order to present more clearly the nature of the interaction. It appears that in Bereiter-Engelmann, DARCEE and controls, females were rated higher than males, whereas in Traditional the reverse was the case. In Montessori there was no consistent difference between the sexes.

Binet-PSI Rater Correlations

Correlations between Binet and PSI testers for each factor on the Face Sheet were based on the spring ratings only. The correlation for each factor was relatively low, ranging from .29 to .34. (Table 15). This low correlation existed for each program and the control group. It seems likely that ratings by the PSI testers are more valid since these testers had an opportunity to observe the children in the process of taking a number of tests, whereas the Binet testers observed them only during the administration of the Binet.

TABLE 14

Means on Stanford-Binet Face Sheet Ratings by Binet and PSI Testers for Programs and Controls

Fall 1968

Program	S-B Testers AC	S-B Testers CN	S-B Testers AL	PSI Testers AC	PSI Testers CN	PSI Testers AL
Bereiter-Engelmann	-	-	-	2.79	2.80	2.72
DARCEE	-	-	-	2.83	2.90	2.88
Montessori	-	-	-	2.98	2.84	2.75
Traditional	-	-	-	2.66	2.66	2.48
Controls	-	-	-	2.82	2.49	2.45
Spring 1969						
Bereiter-Engelmann	2.06	2.13	2.08	2.44	2.36	2.31
DARCEE	2.24	2.10	2.25	2.52	2.51	2.40
Montessori	2.63	2.56	2.61	2.88	2.73	2.56
Traditional	2.30	2.23	2.31	2.72	2.71	2.57
Controls	2.11	2.08	2.08	3.04	2.97	2.86

Note.- Low rating is optimum.

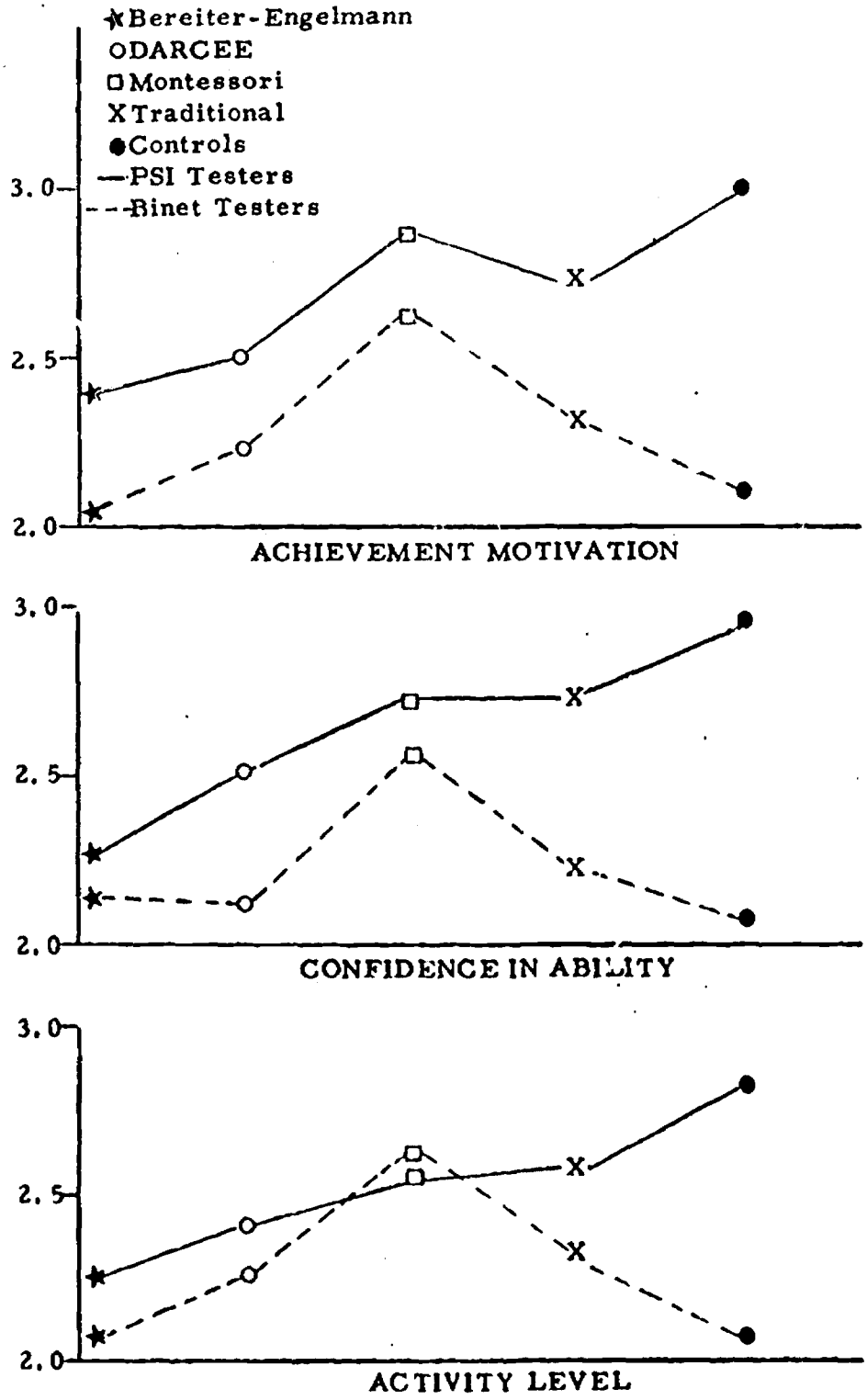

Fig. 22. Spring means on Stanford-Binet Face Sheet ratings by Binet and Preschool Inventory testers for programs and controls. (Low rating is optimum.)

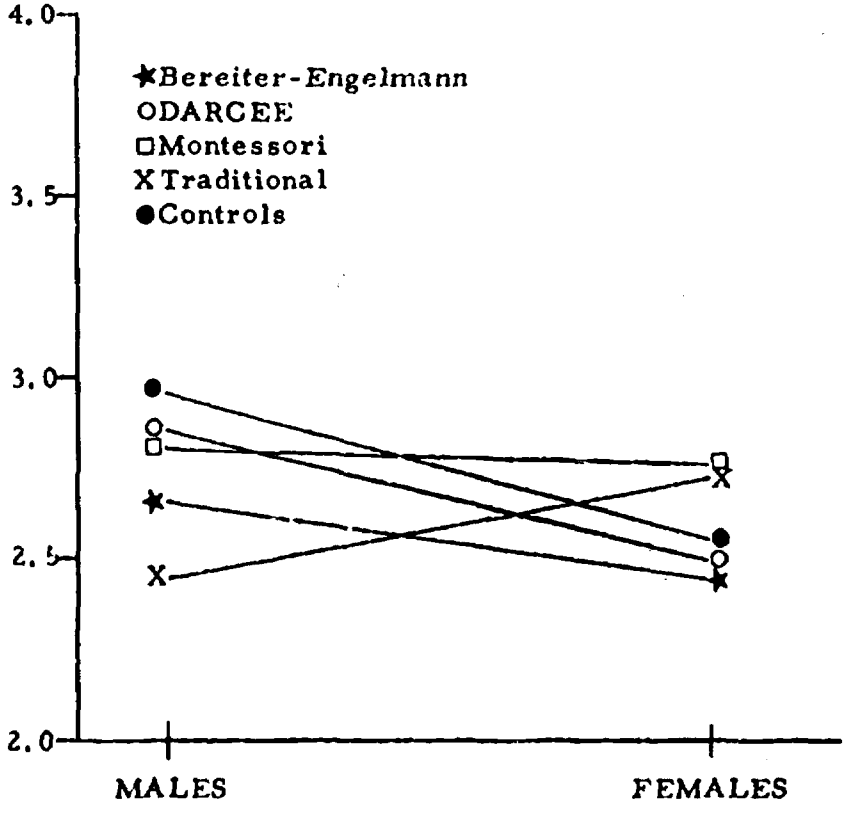

Fig. 23. Male-Female means on combined factors of the Stanford-Binet Face Sheet ratings by Preschool Inventory testers. (Low rating is optimum.)

TABLE 15

Binet-PSI Tester Correlations on Stanford-Binet Face Sheet

		Intercorrelations[a]			PSI with Binet		
		AC	CN	AL	AC	CN	AL
All Programs	AC		.75	.73	.34		
	CN	.84		.90		.29	
	AL	.78	.88				.31
Bereiter-Engelmann	AC		.75	.76	.25		
	CN	.86		.91		.28	
	AL	.79	.89				.32
DARCEE	AC		.68	.58	.42		
	CN	.79		.90		.10	
	AL	.70	.84				.13
Montessori	AC		.77	.67	.38		
	CN	.92		.89		.32	
	AL	.91	.88				.27
Traditional	AC		.59	.72	.33		
	CN	.58		.89		.48	
	AL	.59	.75				.26
Controls	AC		.95	.91	.42		
	CN	.91		.93		.43	
	AL	.85	.94				.56

[a] Correlations for PSI testers above diagonal; Binet testers below.

Summary

Summarizing over motivational and social variables assessed by tests and ratings, analyses suggest that the DARCEE program had considerable impact on children's motivation to achieve, persistence, resistance to distraction, inventiveness, curiosity, independence, timidity, and verbal-social participation. Superiority over controls was found in persistence, motivation to achieve, aggression for Bereiter-Engelmann children; in persistence for Traditional children.

4. Perceptual Measures

(a) Visual

Results from the Early Childhood Embedded Figures Test of the CAB (Table 10) did not reveal any program differences. Children in all programs gained from fall to spring as shown in Figure 24, but controls also improved. One class in the Bereiter-Engelmann program failed to gain, and this produced a pre-post-by-area-by-program interaction. Otherwise, results suggest a maturational process. No sex differences were found. No other measures of perceptual functioning in the visual modality were used.

(b) Auditory

Testing of the original experimental sample with the Wepman Auditory Discrimination Test yielded 55% invalid tests. It appeared doubtful, therefore, whether the remainder of the tests could be considered a valid measure of auditory discrimination for the population of four-year-olds in this experiment. The problem appeared to lie in the administrative format of the Wepma ticularly in the difficulty of the children in und ing "same" and "different". The California Auditor imination Index (CADI), recently developed for us preschool children by Stern (1969), avoids these bies by providing picture choices (half nonsen familiar) which can be selected by pointing on t of labels spoken by the tester.

In order to compare results on the s and obtain a more reliable estimate of audi crimination in the target population, both the Wepm e CADI were administered to the middle-class c roup (N = 48) and to a sample of 48 Head Start childi he same schools the experimental children atte previous year.

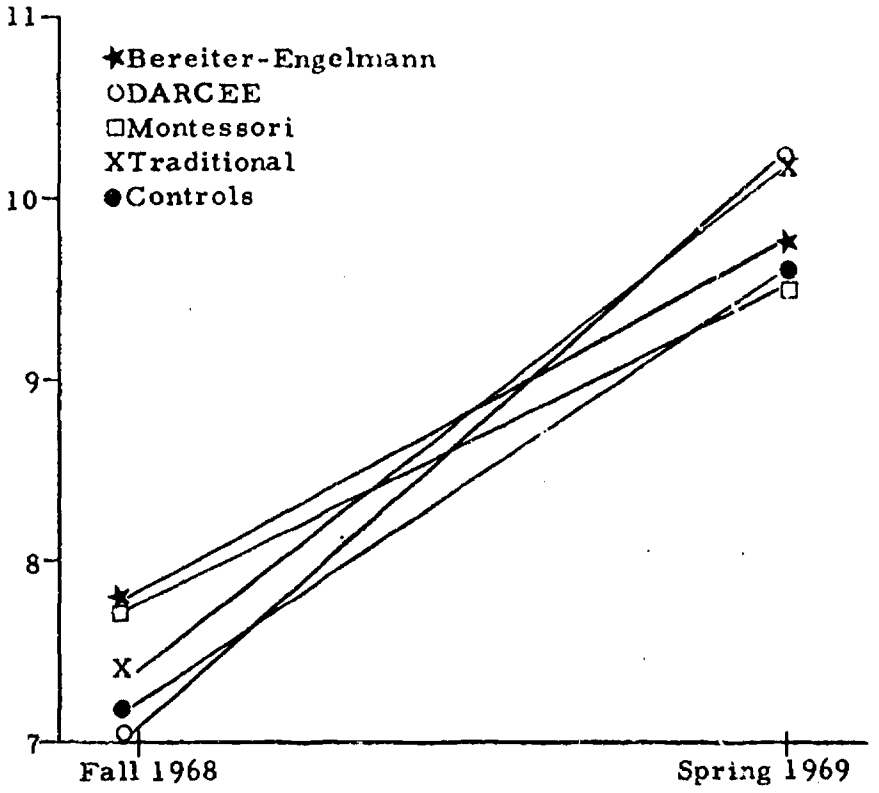

Fig. 24. Fall-Spring means on Embedded Figures for programs and controls.

Instructions and procedure on the Wepman were modified in order to make the test easier to understand. Children were asked to say "same" or "not the same" and additional examples were used. Testers were counterbalanced with respect to the two groups of subjects and order of test administration was balanced by alternation.

Modification of instructions produced valid Wepmans for all but five Head Start and two middle-class children. Table 16 shows error scores for the original experimental sample (Head Start I), the second sample (Head Start II), and the middle-class group. Since the means for the two Head Start groups are the same ($t = .001$, df, 153), it appears that the valid tests on approximately half of the original sample were reliable. The middle-class group was superior, but this is primarily due to the poor performance of Head Start females as compared with middle-class females, as shown in a significant SES-by-sex interaction.

Table 16 also shows the correct response means on the CADI. Again, the middle-class group was superior, but only on nonsense pictures. When the correct response was a familiar figure, there was no difference as a function of SES and both groups were very close to ceiling (19). There was no sex interaction in this test.

Contrary to expectation there was no significant correlation between the two tests for either Head Start (.014) or the middle-class children (-.238).

To determine whether the children had more difficulty discriminating phonemes in initial or final positions, the mean percent passing each type item was computed. Although others (Coller et. al., 1965) have found end sounds more difficult, percent passing final and initial sounds did not differ on the Wepman. On the CADI, however, both Head Start and middle-class groups made significantly more correct responses to items differing in initial phonemes than to those differing in final phonemes.

The lack of correlation and the failure to find end sounds more difficult on the Wepman casts doubt on the Wepman as a measure of auditory discrimination for this population of disadvantaged four-year-olds.

Summary

Programs were not differentiated by the tests of visual and auditory perception used. However, a second study, using both the Wepman and the CADI to measure auditory discrimination indicated that the Head Start females were inferior to middle-class females on the Wepman, and both sexes in Head Start performed poorly on the CADI as compared with the

TABLE 16

Means for Error Score on Wepman and Correct Responses on CADI

	Head Start I	Head Start II	Middle-Class
Wepman	6.46	6.46	4.15
CADI	-	28.96	33.79
Familiar	-	17.67	18.19
Nonsense	-	11.25	15.60

middle-class controls on unfamiliar words. The two tests were virtually uncorrelated for the Head Start groups and not significantly correlated for the middle-class group.

5. Intercorrelations

Both fall and spring intercorrelations for all experimental subjects are given in Table 17. Since similar correlational patterns existed on fall and spring, only the spring correlations are discussed. In addition, the basic correlational patterns for males and females were similar for both testings. Only the spring correlations are presented and discussed (Table 18).

a. Fall-Spring

All fall-spring correlations were significant at the .05 level except for the Resistance to Distraction measure on the Replacement Puzzle (Table 17). The highest correlations over time were for the Preschool Inventory and the Stanford-Binet, .79 and .70 respectively. Fairly high correlations (r = .45 to .60) were found for all factors on the Behavior Inventory as well as for the Activity and Verbal scores on the Curiosity Box. Relatively low correlations (r = .20 to .39) were found on the Face Sheet factors, Dog and Bone, Replacement Puzzle - Resistance, Embedded Figures and the Quick.

The magnitude of these correlations were essentially the same for both males and females (Table 18) with only three exceptions. The correlation for Persistence on the Replacement Puzzle was not significant for females but was .40 for males. The correlation for the Verbal score on the Curiosity Box was .26 for females but .60 for males. Test-retest correlation for Confidence in Ability (Factor II on the Binet Face Sheet) was not significant for either males or females but did reach significance for the total population.

b. Cognitive and Achievement Measures

The correlations among the eight cognitive and achievement measures were quite high. The only non-significant correlation was between Arithmetic and the Peabody Picture Vocabulary Test. In general, the Arithmetic and the Quick tests correlated the lowest with the other tests, the correlations ranging from r = .27 to .54. The correlations among the other remaining tests ranged from .47 to .72. The three measures of intelligence--Stanford-Binet, Quick, and Peabody Picture Vocabulary Test--had moderate intercorrelations(r = .35 to .47).

The strength of the relationship among all the cognitive and achievement measures suggests a general intelligence factor. However, the tendency for Arithmetic to correlate

TABLE 17

Fall-Spring Correlations Among the Dependent Variables for Subjects in the Experimental Programs

	PSI	Q	S-B	PPVT	PSP	EVI	BCI[a]	ARIT	D-B	R-P	R-R	C-A	C-V	EBF	B-TM	B-VP	B-IN	B-AG	B-AC	F-AC[a]	F-CN[a]	F-AL[a]	F-S
PSI		43	61	57	67	71	72	35	23	03	16	18	12	17	34	35	39	11	46	55	57	55	79
Q	31		42	35	40	54	28	28	02	03	16	07	-09	24	17	12	23	13	24	28	27	22	39
S-B	63	42		47	54	49	66	27	13	-02	10	15	15	12	14	25	18	16	22	38	37	34	70
PPVT					59	67	51	09	16	06	07	01	14	33	21	26	34	02	34	36	31	32	
PSP						60	57	41	-05	03	11	17	09	04	12	18	33	24	38	42	34	35	
EVI							67	36	-03	-13	-03	12	08	17	29	29	41	16	34	36	30	33	
BCI[a]								33	20	03	03	11	14	12	25	25	32	24	27	45	40	39	
ARIT									10	-02	20	-03	-08	12	-01	14	-01	03	08	17	04	12	
D-B	18	03	13							-07	-02	14	05	-04	17	10	11	-06	07	08	12	15	31
R-P	14	23	06						05		31	-07	-00	00	-05	-03	04	16	05	22	06	05	20
R-R	04	03	02						-08	15		-03	-02	09	03	10	11	15	07	34	07	05	10
C-A	12	06	13						17	01	-06		26	-14	16	14	07	-10	06	07	27	29	45
C-V	08	-03	06						26	04	05	28		-00	11	13	-09	-10	-04	02	17	30	45
EBF	23	28	14						-04	17	16	09	07		17	09	17	10	19	18	11	10	26
B-TM	22	20	19						22	05	05	26	16	22		56	62	02	70	30	32	36	51
B-VP	26	13	25						23	07	-04	22	27	09	66		39	-16	51	19	24	23	60
B-IN	35	26	26						12	14	08	17	09	38	77	49		20	77	34	28	25	46
B-AC	09	09	10						-08	07	05	00	-09	-05	-10	-23	-09		20	30	07	07	52
B-AC	43	22	36						19	12	07	18	10	20	76	54	76	07		56	53	50	47
F-AC[a]	47	29	38						18	29	14	06	07	24	12	19	17	10	16		70	69	29
F-CN[a]	44	27	31						17	09	09	15	20	11	18	23	16	-10	13	75		89	20
F-AL[a]	47	27	34						23	17	02	24	21	18	24	30	21	-12	22	74	85		30

Notes. - Fall correlations are below the diagonal; Spring above. N=214 for all tests except PPVT, PSP, EVI, BCI, and ARIT where N=84. Correlations significant at $p < .05$ are underlined.

[a] Correlation signs have been reversed for consistency with other instruments.

TABLE 18

Spring Correlations Among the Dependent Variables for Males and Females in the Experimental Programs

	PSI	Q	S-B	PPVT	PSP	EVI	BCI[a]	ARIT	D-B	R-P	R-R	C-A	C-V	EBF	B-TM	B-VP	B-IN	B-AG	B-AC	F-AC[a]	F-CN[a]	F-AL[a]	F-S
PSI		41	64	63	69	73	74	47	11	02	09	10	20	14	32	32	37	12	46	50	57	55	74
Q	45		35	46	44	66	39	30	-13	-08	11	03	-04	20	10	11	08	06	22	30	30	27	42
S-B	58	48		64	66	61	69	31	08	-07	10	22	26	17	12	27	16	15	17	42	50	42	73
PPVT	56	28	36		70	74	49	-01	-15	23	-20	-06	11	27	23	36	47	13	42	44	39	35	
PSP	66	37	46	53		63	81	41	-11	14	11	00	24	19	32	28	46	32	63	52	44	46	
EVI	62	43	38	64	60		86	20	-28	00	-14	07	21	25	43	36	44	13	44	40	37	37	
BCI[a]	71	18	66	55	39	45		51	21	20	31	08	20	06	21	22	36	30	28	42	47	48	
ARIT	35	29	25	22	44	42	23		18	06	08	-19	08	30	32	20	24	22	37	42	43	34	
D-B	35	17	21	37	01	23	20	21		03	02	10	02	-20	11	03	03	05	-06	10	07	11	17
R-P	03	13	03	15	-11	-33	06	03	-16		29	00	06	-00	-06	-09	00	27	08	37	04	04	40
R-R	20	19	09	27	10	04	19	38	-04	14		-04	04	-01	03	06	14	03	11	37	06	07	10
C-A	29	13	10	02	31	18	14	09	16	-15	00		22	-10	11	13	13	-11	02	01	23	24	42
C-V	06	-13	03	06	-02	-06	08	-19	06	-07	-06	29		-10	04	15	-14	-07	-09	06	15	15	60
EBF	20	26	05	37	-03	11	15	04	09	01	16	-19	08		09	09	16	21	20	25	17	14	24
B-TM	35	23	15	25	05	16	20	-20	22	-06	06	21	18	23		55	55	00	66	17	23	22	43
B-VP	37	13	21	26	06	02	29	09	18	-03	13	17	13	10	55		34	-18	49	11	21	22	46
B-IN	41	37	19	33	25	38	26	-16	19	07	09	00	-04	18	62	47		30	24	20	33	31	46
B-AG	10	21	15	-03	16	18	19	-10	-16	04	24	-07	-12	02	04	-15	11		25	32	19	03	37
B-AC	46	25	25	36	22	26	26	-11	18	03	04	00	01	18	73	34	60	15		47	40	41	59
F-AC[a]	60	26	33	35	34	31	47	06	07	06		16	03	10	41	25	27	27	36		62	60	17
F-CN[a]	58	26	29	26	26	25	34	14	16	07	06	31	20	09	38	27	35	05	29	74		67	10
F-AL[a]	56	19	30	30	27	29	32	02	16	06	04	33	25	07	36	24	24	10	26	78	80		23
F-S	80	38	66						44	-03	10	43	26	21	30	54	46	46	52	25	12	21	

Notes. - Correlations for males are above diagonal; females below. For males, N=98, for females N=116 for all tests except the PPVT, PSP, EVI, BCI, and ARIT where N=36 for males, N=48 for females. Correlations significant at $p < .05$ are underlined.

[a] Correlation signs have been reversed for consistency with other instruments.

at a lower level with the other tests implies both verbal and quantitative dimensions on this factor.

This pattern of high correlations also existed for both males and females. However, the intercorrelations tended to be slightly higher for males than for females, especially for the Basic Concept Inventory (BCI). For females the BCI did not correlate significantly with Arithmetic or the Quick, only .39 with Parallel Sentence Production (PSP), and .45 with Expressive Vocabulary Inventory (EVI). However, for males the correlation with the BCI was .51 for Arithmetic, .39 for the Quick, .81 for PSP and .86 for EVI. Two similar situations existed for males and females on the Stanford-Binet. For females the Binet correlated .38 with both the Peabody Picture Vocabulary Test and EVI, while for males the correlation with the Peabody was .64 and was .61 with the EVI.

c. Social and Motivational Ratings

All the correlations between teacher ratings on the Behavior Inventory were significant except that between Aggression and Timidity. In general, Aggression correlated the lowest with the other factors ($r = .16$ to $.20$), while the correlations among the remaining factors were generally high ($r = .39$ to $.77$). Three factors--Independence, Achievement, and Timidity--were highly interrelated, with the correlations ranging from .62 to .77. The correlation between the Achievement Motivation measures on the Behavior Inventory and Face Sheet was .36. In general, correlations between the Face Sheet and Behavior Inventory factors were significant but low, ranging from .19 to .36.

The same general pattern existed for males and females. However, on the Behavior Inventory some slight variations occurred. For females, Aggression was not significantly correlated with any of the factors. However for males Aggression correlated negatively with Verbal-Social Participation and positively with Independence and Achievement ($r = -.18$, $.30$, and $.25$, respectively). Independence, Achievement, and Timidity clustered for both males and females, but the cluster was stronger for females.

d. Motivational and Perceptual Tests

Very few relationships existed among the motivational and perceptual tests: Dog and Bone, Replacement Puzzle, Curiosity Box, and Embedded Figures. The two relationships that existed for all experimental subjects, as well as for both males and females, were those between the two scores of the Replacement Puzzle and the two scores of the Curiosity Box, with the average correlation being approximately .30 for both tests.

e. Cognitive and Achievement Tests with Social and Motivational Ratings

In general, the correlations between the cognitive tests and the rating scales were moderate, the majority being between .20 and .45. The exception to this pattern was the lack of relationship between Arithmetic and the motivational ratings.

Moderate correlations also existed for both males and females. However, three major differences occurred. For females, Arithmetic did not correlate with any of the ratings. However, for males Arithmetic did correlate with each of the Face Sheet factors and with Timidity and Achievement Motivation on the Behavior Inventory. Aggression did not correlate with any of the achievement tests for females, but did correlate with the PSP and the BCI for males. For females, the Quick correlated with all the Behavior Inventory factors except Verbal-Social Participation, but for males the Quick correlated only with the Achievement Motivation factor.

f. Cognitive and Achievement Tests with Motivational and Perceptual Tests

In contrast with the moderate correlations between the cognitive measures and the behavioral ratings, there was little relationship between the cognitive and motivational tests, indicating the independence of these two sets of measures. Since different correlational patterns were found for males and females, the relationship between these two sets of variables will be described by sex rather than for the total population.

In general, more significant correlations between the cognitive and motivational tests occurred for females ($r = .17$ to $.37$) than for males ($r = .22$ to $.31$). The major distinctions are as follows. For females, Dog and Bone and Replacement Puzzle - Resistance correlated with the Preschool Inventory, Peabody and the Quick, while none of these relationships existed for males. For females, Resistance to Distraction also correlated with Arithmetic, and the Dog and Bone correlated with the Stanford-Binet. For males the only variables that correlated with the Binet were the Activity and Verbal scores on the Curiosity Box. For females, Activity on the Curiosity Box correlated with the Preschool Inventory and PSP. Three negative relationships occurred. For females the EVI correlated negatively with Persistence; however, for males it correlated negatively with Dog and Bone. Also, for males, the BCI correlated negatively with Resistance.

For the Embedded Figures Test, four correlations were significant for males, two for females. The only agreement between these two patterns was the significant correlation with the Quick (r = .20 and .26).

It is extremely difficult to interpret the meaning of each of these correlations, since the absolute size is small. However, the distinct difference between sexes in the correlational pattern is significant.

g. Social and Motivational Ratings with Motivational and Perceptual Tests

Correlations between the social-motivational ratings and tests were generally low (r = .12 to .34) or not significant. More significant correlations occurred for females than males. One example of this difference is that for females the Dog and Bone correlated with all factors of the Behavior Inventory except Aggression, while for the males the Dog and Bone did not correlate with any Behavior Inventory factors.

For both males and females the Embedded Figures Test correlated with Achievement Motivation on the Behavior Inventory. However, this was the only consistency between the two correlational patterns.

As with the cognitive and motivational tests, there was little relationship between the motivational ratings and the motivational tests. Again, different patterns were found for males and females.

Summary

The fall-spring correlations for the dependent variables were generally moderate to high. The intercorrelation patterns among these variables were similar for both fall and spring. Variables may be grouped into three categories: the achievement and cognitive measures which were highly interrelated, the social and motivational ratings which were moderately interrelated, and the motivational tests which were not interrelated. Similar intercorrelations were found for both males and females for each of these categories. Cognitive measures and the social-motivational ratings both correlated poorly with the motivational tests. However, different patterns for males and females occurred. Correlations between the cognitive measures and social-motivational ratings were moderate and indicated no sex differences.

Within Programs

(1) Class Differences

Teacher or class differences emerged in this design as area-by-program interactions. Although they were not large enough in most cases to overshadow the effects of programs, a number of them did occur. No systematic attempt has been made as yet to interpret them.

(2) DARCEE Home Visitor Study

Within the DARCEE program the mothers of approximately one-half of the children were visited once a week in their homes. Two Home Visitors were employed, and these individuals attended the 8-week training program with the teachers. All material used in the classrooms were supplied to them and they attempted to give the mothers an understanding of the curriculum, its goals and methods. They encouraged the mothers to use the materials in the home and helped them devise other materials which would be of value for their children. Each Home Visitor was attached to two classes and called on about half the mothers in each class.

Visited and not-visited children were compared on all dependent variables. No significant differences existed between the two groups.

F. Relations Between Treatment Dimensions and Treatment Effects

As a preliminary analysis, multiple regressions were done to assess the relationships between teaching techniques obtained from the in-class monitoring procedure and change scores (fall-spring) on each of the dependent variables. Scores on each of the dependent variables were available for each subject; however, scores on teaching techniques were available only for classes.

The program variable was not included in the regression analysis since the purpose was to examine the relationship between the activities assessed by in-class monitoring and changes in the dependent variables independent of knowledge of the program itself. It was already established that programs differed on the in-class monitoring variables and also on the dependent variables. Thus, the effect of including program in the regression analysis would be to attenuate the beta weights for those classroom variables highly correlated with programs.

Since the values of the predictor variables were means for classes and the values of the dependent variables were individual scores, high multiple correlation coefficients were not expected. If the variability of change scores within each of the classes was small, then the multiple correlation could be expected to be high. However, if the variability of the change scores within the classes was great, (which might be the

case, for example, if the teacher paid more attention to some children than others), then the multiple correlations between the in-class variables and the dependent variables would be an underestimation of the actual relationship.

Multiple correlations for the predictor variables ranged from .23 to .42. Partial correlations between the criterion variables and the given predictor, holding the other predictors constant, ranged from -.29 to .31.

In view of the significant differences among programs on videotape variables such as reinforcement, the inclusion of these variables would be desirable. The large number of potential predictors, however, makes selectivity essential. Sex differences on some dependent measures indicate that regression analyses should be made separately for the sexes. Finally, a decision regarding whether to predict final level or change must be made on both logical and empirical grounds for each variable.

VI. SUMMARY AND DISCUSSION

The major results of this experiment provide answers to a number of questions which have been previously raised with regard to preschool programs.

1. The four programs compared were found [illegible]r significantly along a number of dimensions. The question of [illegible] various programs differ in actual operation as well as descriptively can only receive a partial answer by means of consultants' evaluations. Although the implementations in this study were not prototypes of the originals, consultants considered them to be moderately successful. More important, however, is the fact that the four programs were shown to differ along a number of specific dimensions, and in predicted directions.

With samples of four teachers in three of the programs and two in the fourth, program differences emerged clearly, despite within-program teacher differences on the variables assessed by monitoring procedures. In respect to such individual difference variables as intelligence, personality, and experience, there was wide variation among the 14 teachers, but the differences were not related to which program the teachers were using. There is no evidence, therefore, from this study that individual teachers' characteristics were a source of differences among programs. This suggests that regardless of background and individuality, preschool teachers can be greatly influenced in their teaching methods by brief training sessions which involve specific and unambiguous instruction regarding techniques. Results also indicate that four to eight weeks of such teacher training, plus two on-site visits by consultants, can suffice for identifiable program implementations, even in the absence of the expertise, dedication, and enthusiasm of program developers.

2. Programs had significantly different effects on children. The immediate effects of two of the programs on children, although not so dramatic as those which have sometimes been found by program developers, were statistically significant, in predictable areas of development, and large enough to appear psychologically meaningful. The effects of the Bereiter-Engelmann program were largely confined to cognitive and academic areas, a result consistent with the highly focused nature of the program. The effects of the DARCEE program were more diffuse, and most evident in the areas of motivation and attitudes. The total ineffectiveness of the Home Visitor Program in DARCEE was a somewhat unexpected result. However, Miller (personal communication) found the effects of a Home Visitor program to be greater on younger siblings of the children in preschool than on the target children. It may also be the case that the continuing effectiveness of this program as long as six years after two summer programs (1968) was a function of the home visitation. If this is the case, it may be possible to detect the effects at subsequent retesting.

A number of sex effects occurred, but these are difficult to interpret or summarize.

In general, results indicate that the immediate impact was superior for the two programs which are most didactic.

REFERENCES

Ammons, R.B., and Ammons, C.H. "The Quick Test (QT)". Provisional Manual. Psychological Reports, (Monograph Supplement), 1:111-161, 1962.

Bales, Robert F. Interaction Process Analyses. Cambridge, Mass.: Addison-Wesley Press, 1950.

Banta, Thomas J. "Tests for the Evaluation of Early Childhood Education: The Cincinnati Autonomy Test Battery (CATB)" Cognitive Studies, Vol. 1, 1968.

Bellar, E.K. "The Evaluation of Effects of Early Educational Intervention on Intellectual and Social Development of Lower-Class, Disadvantaged Children", Critical Issues in Research Related to Disadvantaged Children, Edith Grotberg, Ed., Princeton, New Jersey: Educational Testing Service, 1969, Seminar #3.

Biddle, B.J., and Ellena, W.J., Eds. Contemporary Research in Teacher Effectiveness. New York: Holt, Rinehart, and Winston, 1964.

Coller, A., Schwartz, S., & Coleman, R. "Auditory Discrimination Tests". Paper prepared for the Institute of Developmental Studies, New York, 1965.

Cupp, Janet R. "An Instructional Program for Three- and Four-Year-Old Disadvantaged Children". Demonstration and Research Center for Early Education, George Peabody College, Nashville, 1967.

Elashoff, Janet D. "Analysis of Covariance: A Delicate Instrument", American Educational Research Journal, Vol. 6, 1969, 383-402.

Gagne, R.M. "Contributions of Learning to Human Development", Psychological Review, 75(3), 1968, 177-191.

Gordon, I. Discussant on Siminar #2, in Critical Issues in Research Related to Disadvantaged Children, Edith Grotberg, Ed., Princeton, New Jersey: Educational Testing Service, 1969.

Gray, S.W., and Klaus, R.A. "The Early Training Project for Disadvantaged Children: A Report After Five Years", Monographs of the Society for Research in Child Development, Vol. 33, No. 4, 1968.

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

Hess, R.D., Kramer, D., Slaughter, D., Torney, J., Berry, D., and Hull, E. "Techniques for Assessing Cognitive and Social Abilities of Children and Parents in Project Head Start", Report on Research Contract OEO-519 with Office of Economic Opportunity, University of Chicago, July, 1966.

Hodges, W.L., and Spicker, H.H., "The Effects of Preschool Experiences on Culturally Deprived Children" in The Young Child, W.W. Hartup and N.L. Smothergill (Eds.), Washington, D.C.: National Association for the Education of Young Children, 1967, 263-289.

Hymes, James, L. Jr., Teaching the Child Under Six. Columbus, Ohio: Charles E. Merrill Publishing Company, 1968.

Hunt, J. McV., Intelligence and Experience. New York: The Ronald Press Company, 1961.

Jensen, A.R., "How Much Can We Boost I.Q. and Scholastic Achievement?" Harvard Educational Review, 1969, 39, 1-123.

Kagan, Jerome. "The Concept of Identification", Psychological Review, Vol. 65, No. 5, 1958.

Karnes, M.B., Taska, J.A., and Hodgins, S.A., "A longitudinal study of disadvantaged children who participated in three different preschool programs". Paper presented at the annual meeting of the American Educational Research Association, Los Angeles, Calif., 1969.

Kirk, Roger E. Experimental Design: Procedures for the Behavioral Science. Belmont California: Brooks/Cole, 1968.

Kirk, S.A., Early Education of the Mentally Retarded, Urbana, Illinois: University of Illinois Press, 1958.

Kounin, J.S., "An Analysis of Teachers' Managerial Techniques", Psychology in the Schools, Vol. 4, 1967, 221-227.

Levin, Joe. R., & Marascuilo, Leonard A. "Appropriate Post Hoc Comparisons for Interaction and Nested Hypotheses in Analysis of Variance Designs: The Elimination of Type IV Errors", American Educational Research Journal, 1970, 7(3), 397-421.

Miller, J.A., "Intervention Research with Young Disadvantaged Children and their Parents", Paper presented to the American Psychological Association, Washington, D.C., 1967.

Miller, N.E., & Dollard, J., Social Learning and Imitation. New Haven: Yale University Press, 1941.

Neill, Robert D., & Bein, Sara Lee. "An Attitude Scale for Teachers of the Culturally Disadvantaged", Tehcnical Supplement No.2, Head Start Louisville & Jefferson County Community Action Commission Evaluation, University of Louisville, Louisville, Kentucky, June 9, 1967.

Rainbow Series, Project Head Start, Office of Economic Opportunity, Washington, D.C., 1965.

Rashid, M. "The Teacher, Teacher Style, and Classroom Management" in Critical Issues in Research Related to Disadvantaged Children, Edith Grotberg, Ed., Princeton, New Jersey: Educational Testing Service, 1969, Seminar #2.

Sears, P.S., and Dowley, E.M., "Research on Teaching in the Nursery School", Handbook on Research on Teaching, N.L. Gage, Ed., Chicago, Illinois: Rand McNally, 1963, 814-864.

Scheffe, Henry. "A Method for Judging all Contrasts in the Analysis of Variance". Biometrika, 40 (1953), 87-104.

Sprigle, H.A., Van de Riet, and Van de Riet, H. "A Sequential Learning Program for Preschool Children and an Evaluation of its effectiveness with Culturally Disadvantaged Children", Paper presented at the American Educational Research Association, New York, March, 1967.

Sprott, D.A. "Note on Evans and Anastasio on the Analysis of Covariance", Psychological Bulletin, 1970 (73), 303-306.

Stern, Carolyn. "Children's Auditory Discrimination Inventory (CADI)", Mimeograph, University of California, June, 1969.

Weikart, D.P., "A Comparative Study of Three Preschool Curricula", Paper presented at the ibennial meeting of the Society for Research in Child Development, Santa Monica, Calif., March, 1969.

Winer, B.J. Statistical Principles in Experimental Design. New York: McGraw-Hill, 1962.